Racial Discrimination in Federal Sentencing: Evidence from Drug Mandatory Minimums*

Cody Tuttle

September 17th, 2018

This is a preliminary draft.

Abstract

I test for racial discrimination in the criminal justice system by analyzing abnormal bunching in the distribution of crack-cocaine amounts used in federal sentencing. I compare cases sentenced before and after the Fair Sentencing Act, a 2010 law that changed the 10-year mandatory minimum threshold for crack-cocaine from 50g to 280g. Using data at multiple stages in the criminal justice process, I find the following: (1) after 2010, there is a sharp increase in the fraction of cases at 280g, the amount that now triggers the 10-year mandatory minimum; (2) this increase is disproportionately large for black and Hispanic offenders; (3) this increase is driven by prosecutors; (4) the fraction of cases at 280g falls once evidentiary standards become stricter; and (5) the racial disparity in the increase cannot be explained by differences in education, sex, age, criminal history, seized drug amount, or other elements of the crime, but it can be almost entirely explained by a measure of state-level racial animus. These results shed light on the role of prosecutorial discretion and racial discrimination as causes of racial disparities in sentencing.

JEL Classification: J15, K14, K41, K42

Keywords: Racial discrimination, mandatory minimums, prosecutors, sentencing, bunching

*University of Maryland, College Park, Department of Economics, 3114 Tydings Hall, College Park, MD 20740. Email: tuttle@econ.umd.edu. For helpful comments, I thank Prateik Dalmia, Judy Hellerstein, Brian Johnson, Ethan Kaplan, Melissa Kearney, John Laub, Nolan Pope, Matthew Staiger, Lesley Turner, Riley Wilson, and the participants of the University of Maryland Microeconomics Workshop and the University of Maryland Political Economy Workshop. I thank Stephanie Rennane for sharing her bunching analysis code. Finally, I am grateful to Robert Breetz and several anonymous attorneys for feedback and insights about the institutional setting.

I. Introduction

Racial differences in sentencing are a persistent problem in America. In recent federal cases, black offenders face sentences that are 20 percent longer than the sentences handed down for white offenders (United States Sentencing Commission (USSC) 2017).[1] These added years are costly for society at large and for the people incarcerated. The Bureau of Prisons (BOP) estimates the marginal cost of incarcerating an additional person is about $11,000 (in 2015 dollars) per year (Department of Justice (DOJ) 2011). Furthermore, Mueller-Smith (2015) estimates an additional year in prison causes a 30 percent decrease in formal earnings post-release and significant lost wages while incarcerated. Due to racial sentencing disparities, these economic costs are disproportionately borne by black and Hispanic offenders.[2] Of course, for policy to confront these disparities, we must understand the root causes. One explanation for disparate sentences is that defendants of different races are different *upon entry* into the criminal justice system. Another explanation, however, is that *after entry* into the system, defendants are treated differently by race.

In this paper, I test the second explanation, that agents in the criminal justice system (police, prosecutors, judges, etc.) treat black and Hispanic defendants differently than similar white defendants. To do this, I focus on federal crack-cocaine cases and the application of mandatory minimum sentences. Approximately 20 percent of all federal cases involve a crack-cocaine offense, and racial sentencing differences are particularly large in these cases. In 2016, black and Hispanic crack-cocaine offenders were sentenced to over 6 years, on average, compared to only 3.5 years for white crack-cocaine offenders (USSC 2017). In addition, the structure of mandatory minimum sentencing and recent changes in crack-cocaine mandatory minimums provide a unique opportunity to study discretion and discrimination in the criminal justice system.

In federal drug trafficking cases, a mandatory minimum sentence is triggered if the drug trafficking crime involves an amount of drugs equal to or above a threshold amount. This sentencing cliff generates strong incentives for law enforcement agents. Furthermore, legal rules about police sting operations and the type of evidence admissable in federal court give both police and prosecutors power to manipulate the amount used in sentencing. If police or prosecutors want to increase the likelihood of a harsh sentence, they can use their discretion to report an amount of drugs at or just above the threshold amount. I employ a bunching estimation design to determine whether police or prosecutors respond to this sentencing incentive and whether their responses reflect racial discrimination. Specifically, I test for excess mass at and above the mandatory minimum threshold (i.e. the use of discretion to increase the likelihood of a harsh sentence)

[1] In a 2017 report, the USSC also notes that Hispanic men are given sentences that are 5.3 percent longer than sentences for white men. As discussed below, the Hispanic-white sentencing disparity is much larger in crack-cocaine cases, the primary focus of this paper.

[2] In the United States Sentencing Commission (USSC) data files, four values are recorded for the offender's "race"–(1) white, (2) black, (3) Hispanic, and (4) other. As such, throughout the paper, I will frequently use the term "race" in reference to black and Hispanic people to be consistent with the terminology used by the USSC.

and for differences in the excess mass by race (i.e. racial bias in the use of discretion).

With the Fair Sentencing Act in 2010, the 10-year mandatory minimum threshold for crack-cocaine was increased from 50g to 280g.[3] Crack-cocaine is the only drug for which the federal mandatory minimum threshold has changed since the adoption of mandatory minimums in the 1980s. The shift to 280g is especially useful since the new threshold is set at a point with zero bunching prior to 2010. All other mandatory minimum thresholds are set at somewhat natural bunching points (50g, 500g, 1000g) that do not vary over time.[4] Using this time variation in the mandatory minimum threshold, I implement a difference-in-bunching design where I assume the pre-2010 distribution of drug amounts is a good counterfactual for the post-2010 distribution (i.e. what the post-2010 distribution would look like in the absence of a 280g threshold) (Kleven 2016). I find the fraction of cases bunched at and above 280g increases after 2010, and that the increase is much larger for black and Hispanic offenders than for white offenders.[5]

To be clear, this is not intended as an evaluation of the Fair Sentencing Act, which is likely responsible for a decline in sentences after 2010 (USSC 2015a). Rather, these results imply that police or prosecutors dampened the effect of the Fair Sentencing Act by increasing the drug amount charged for some offenders. In addition, these results do not imply that the use of discretion or racial bias in the use of discretion began after 2010. Instead, I take the shift to 280g as an opportunity to detect these behaviors that are otherwise difficult to detect.

I use data at multiple stages in the criminal justice process to determine who is responsible for the bunching at 280g. First, since the bunching occurs in federal sentencing, it is possible that more cases with drug quantities at or above 280g are sent to federal court after 2010. I examine data on state-level drug convictions from Florida and North Carolina, and I do not find a shifting composition of cases after 2010. Second, local and federal law enforcement can influence the drug quantity involved in an offense by manipulating amounts involved in sting operations.[6] Data on drug arrests and seizures made by local and federal law enforcement do not show increased bunching at 280g after 2010. Finally, prosecutors can influence the drug quantity involved in an offense because, according to the USSC Federal Sentencing Guidelines, the quantity of drugs used to determine sentencing is not strictly tied to the quantity found on the offender at

[3]I focus on the higher, 10-year mandatory minimum threshold for drugs in this paper. I do not study bunching at 28g of crack-cocaine (the lower, 5-year mandatory minimum threshold) in detail because this amount is below the pre-2010 higher, 10-year mandatory minimum threshold of 50g and thus, offenders in the range from 28-50g may be responding to those shifting incentives. In other words, an offender would face 5 years for holding 29g pre-2010 and 10 years for holding 51g pre-2010. Post-2010, a person holding either of those amounts will face only 5 years. This shift in incentives does not occur for the 280g threshold. In addition, crack-cocaine is often transacted in ounces at low weights and the 28-29g range includes a commonly sold amount: one ounce. However, in Table 3, I do show the change in probability of being recorded with 28-50g (not including 50g) for white and black and Hispanic offenders after 2010. I find that black and Hispanic offenders are slightly more likely to be recorded with 28-50g after 2010.

[4]These amounts exhibit bunching in all drug-types, even those where they are not the relevant thresholds. I expect this bunching is due to: (1) a "round number" effect and (2) actual drug trafficking behavior. Large quantities of drugs are frequently distributed in kilograms (US DOJ 2002)

[5]Note, I do not find evidence of bunching just below 280g for the drug amount used in sentencing.

[6]In sting "sells", undercover law enforcement sell drugs to a suspect, and in those cases, can control the amount sold in order to trigger the mandatory minimum. In sting "buys", undercover law enforcement buy drugs from a suspect, and in those cases, can control the length of the investigation until the amount transacted triggers the mandatory minimum.

the time of arrest (USSC 2015b).[7] I do find bunching at 280g after 2010 in case management data from the Executive Office of the US Attorney. Moreover, I find that approximately 30% of prosecutors are responsible for the rise in cases with 280g after 2010, and that there is variation in prosecutor-level bunching both within and between districts. This suggests that the observed bunching is due to prosecutorial discretion.

Furthermore, the US Supreme Court issued a 5-4 decision in *Alleyne v. United States* on June 17, 2013 that changed the evidentiary standard necessary for facts that raise a defendant's exposure to mandatory minimum sentencing (Bala 2015). Previously, prosecutors could present evidence on drug quantities to the presiding judge, and the judge would decide, based on the preponderance of evidence, whether the mandatory minimum applied. The Supreme Court ruling in *Alleyne* requires that prosecutors present this evidence to the jury which evaluates it based on the stricter "beyond a reasonable doubt" standard. The case management data from the Executive Office of the US Attorney show that from 2011-2013, approximately 9.1% of cases were recorded in the range of 280-290g. From 2014-2016, however, 6.8% of cases were recorded in the 280-290g range. Using a difference-in-discontinuities design, I show that the practice of bunching ballooned in the run up to *Alleyne*, and that this bunching was reigned in by the Supreme Court decision (though it was not eliminated entirely). This suggests prosecutors were submitting evidence under the judicial fact-finding system that would not hold up under the scrutiny of a jury.

I also explore the possibility that the racial differences in bunching at 280g are not driven by race but are instead driven by another factor correlated with race. First, I highlight the fact that the distributions of seized drugs by race are similar both before and after 2010. It does not appear that different trafficking intensity by race or different offender responses to the Fair Sentencing Act are responsible for the racial differences in bunching. Next, I show that racial differences in bunching exist even among observably similar offenders. For example, the increase in cases at and above 280g for black and Hispanic offenders with a college education is larger than the increase for white offenders with a college education. This is also true for interactions by race and sex, age range, criminal history, and other elements of the current offense. Race is a consistent factor in determining the amount of bunching at 280g after 2010. Finally, I show that the racial disparity in bunching can be almost entirely explained by a measure of state-level racial animus based on Google search data developed by Stephens-Davidowitz (2014). In other words, black and Hispanic offenders convicted in states with high levels of racial animus are more likely to be bunched at 280g than white offenders convicted in those states. In states with low levels of racial animus, however, black, Hispanic, and white offenders are all equally likely to be bunched at 280g.

Taken together, these results suggest prosecutors use their discretion to tag some offenders with drug amounts that will trigger mandatory minimum sentences, and that they do this disproportionately for black

[7] The USSC Federal Sentencing Guidelines (2015b) specifically state, "Types and quantities of drugs not specified in the count of conviction may be considered in determining the offense level. Where there is no drug seizure or the amount seized does not reflect the scale of the offense, the court shall approximate the quantity of the controlled substance. In making this determination, the court may consider, for example, the price generally obtained for the controlled substance, financial or other records, similar transactions in controlled substances by the defendant, and the size or capability of any laboratory involved."

and Hispanic offenders. Even more, the decrease in bunching after the Supreme Court tightens evidentiary standards in *Alleyne* suggests these cases are reliant on relatively weak evidence. Finally, the persistent racial differences even after controlling for and interacting race with observables, the within-district variation in prosecutor-level bunching, and the correlation between the racial disparity in bunching and state-level racial animus all support a model of discrimination in which the disproportionate manipulation is a result of prosecutor tastes. Of course, a complicated model of statistical discrimination could incorporate those facts, and I cannot reject such a model. Instead, this paper provides evidence that black and Hispanic crack-cocaine offenders are treated differently than similar white offenders after entry into the criminal justice system, and in this case, prosecutors are likely responsible for the discriminatory treatment.

This paper contributes to the empirical literature on prosecutorial discretion and decision-making (Glaeser, Kessler and Piehl 2000; Bjerk 2005; Boylan 2005; Shermer and Johnson 2010; Rehavi and Starr 2014; Yang 2017; Nyhan and Rehavi 2017). Bjerk (2005), for example, shows that prosecutors are more likely to charge defendants with a misdemeanor if a felony charge would invoke a "three-strikes" sentence. Shermer and Johnson (2010) find that male defendants are less likely to receive a charge reduction than female defendants, but that there are no differences by race or ethnicity. Rehavi and Starr (2014), on the other hand, find that black offenders receive harsher sentences than white offenders arrested for the same crime. Using linked data from US Marshalls, US courts, and US federal sentencing, they show that this disparity is driven by prosecutorial discretion over initial charging decisions. In this paper, I provide additional evidence that prosecutors are selectively harsh (or lenient) by race using a new source of identification. In addition, I quantify the fraction of prosecutors exercising this type of discretion, and I show that this can be mitigated by increasing evidentiary standards.

More broadly, this paper adds to an extensive literature on racial discrimination in the criminal justice system (e.g. Knowles, Persico, and Todd 2001; Anwar and Fang 2006; Grogger and Ridgeway 2006; Antonovics and Knight 2009; Anwar, Bayer, Hjalmarsson 2012; Rehavi and Starr 2014; Agan and Starr 2018; Arnold, Dobbie, and Yang 2018; West 2018). The vast majority of papers on this topic focus on racial bias from police officers, and test for bias using the outcome (or hit-rate) test proposed by Becker (1957) or by documenting same-race versus other-race bias.

Along with the recent work by Anbarci and Lee (2014) and Goncalves and Mello (2018), I implement a new test for racial bias in criminal justice that uses insights from the bunching literature.[8,9] Both Anbarci and

[8]Note, my paper is not the first to acknowledge the existence of bunching in the amount of drugs recorded in US federal sentencing or the possibility that it could be used as a test of prosecutorial discretion and discrimination. However, this paper is the first, to my knowledge, to take advantage of the time variation in the crack-cocaine 10-year mandatory minimum threshold to isolate bunching that is solely due to prosecutor manipulation. In addition, I examine data at multiple stages in the criminal justice process and conduct several additional empirical tests that all suggest bunching is due to prosecutorial discretion and negatively affects minority defendants. Related work in this area is discussed in more detail in Section I.

[9]Recently, economists have also studied cliffs and notches in other settings. For example, Diamond and Persson (2017) and Dee et al. (2017) both find teachers manipulate test scores on high-stakes tests in response to grade cutoffs. Both papers show evidence of substantial bunching right above important grade thresholds, and identify the types of students who are most likely to have their grade manipulated. Diamond and Persson, using data from math tests in Sweden, find that teachers manipulate scores for students

Lee (2014) and Goncalves and Mello (2018) study the prevalence of police officers discounting speeding tickets by race. They show substantial bunching just below the point where the fine increases, and they argue that this is a result of officer leniency. Anbarci and Lee (2014) show that white officers discount more for white drivers and black officers discount more for black drivers. Goncalves and Mello (2018) demonstrate that only some officers practice this leniency and that those officers are, on average, more lenient toward white drivers than minority drivers. I contribute to this literature by examining racial bias from prosecutors (a relatively understudied group), and by showing racial differences in bunching at the point where sentences increase.

Finally, the bunching in drug weights and the racial discrimination in bunching has meaningful sentencing consequences and implications for the racial sentencing gap. Depending on the counterfactual sentence imputed for the affected offenders, bunching at 280g can account for 2-7 percent of the racial disparity in crack-cocaine sentences. A highly conservative estimate suggests that being bunched at 280g adds 1-2 years to an offender's sentence. Multiple estimates suggest the cost of incarceration (combining direct care costs and the cost of lost current and future wages for the offender) is approximately \$60,000 per person per year (Donohue and Sieglman 1998; Donohue 2009; Mueller-Smith 2015). Applying this cost to the 3.5% of crack-cocaine cases bunched at 280g from 2011-2015 implies a total cost of \$16-\$32 million. Assuming 3.5% of all drug cases from 1999-2015 were subject to similar discretion further implies a total cost of \$1-\$2 billion.

All of the calculations above are based on the amount of discretion and discrimination detected right at and above the 10-year mandatory minimum threshold for crack-cocaine. To the extent that prosecutors exercise similar discretion to push offenders just above 5-year mandatory minimum thresholds or exercise discretion in less obvious ways (pushing offenders far beyond thresholds, for example), the cost estimates will only be higher and the effect on racial sentencing differences will only be greater.

This paper proceeds as follows. I detail the institutional background and introduce a simple conceptual model of prosecutor objectives in Section II. In Section III, I describe the various data sources I use. I introduce the methodology in Section IV, and I discuss results in Section V. Section VI concludes.

who perform unusually poorly (given past performance) on the exam. They do not find evidence of discrimination by gender or immigrant status. Dee et al., on the other hand, show that teachers in New York City inflate scores for students near the threshold and that, conditional on being near the threshold, teachers are less likely to manipulate scores for black or Hispanic students. Both of these papers also estimate the consequences of test score manipulation for those students who have their scores inflated. In general, the test score manipulation leads to positive educational outcomes for those students who are bumped above the threshold.

II. Institutional Background and Prosecutor Objectives

A. Institutional Background

Debate about federal mandatory minimum policy has overwhelmingly focused on the disparity between the threshold amounts for crack-cocaine and powder-cocaine. Prior to 2010, the threshold for the crack-cocaine 10-year mandatory minimum was 50 grams whereas the 10-year threshold amount for powder-cocaine was 5000g. In part due to the recommendations of the United States Sentencing Commission, the threshold amounts for crack-cocaine were increased in 2010 by the Fair Sentencing Act. The upper threshold was changed from 50g to 280g. It is not clear why 280g in particular was chosen. One potential reason is that lawmakers wanted to set the threshold at 10 ounces, but in keeping with the convention of setting the threshold in grams or kilograms, chose 280g as the closest "round" number to 10 ounces. In this paper, I use this change from 50g to 280g to study bunching at mandatory minimum thresholds and its relation to discretion and racial discrimination in the criminal justice system.

This paper is not the first to acknowledge bunching in the amount of drugs recorded in US federal sentencing.[10] Bjerk (2017) briefly discusses bunching in the distribution of drug amounts, but posits that bunching arises from negotiation downward by prosecutors and defendants.[11] In addition, a 2015 Bureau of Justice Statistics (BJS) working paper on federal sentencing disparities advances the idea that prosecutors could "game" the drug weight sentencing guidelines (Rhodes, Kling, Luallen, and Dyous 2015). That paper provides a cursory look at bunching above mandatory minimum thresholds for all drugs by race, but does not address the bunching that is always present at round-number amounts (50g, 100g, 500g, 1000g, 5000g). As such, the authors conclude prosecutorial discretion does not differentially affect black and Hispanic offenders.

I depart from previous work in several ways. First, I show that excess mass at the threshold comes from cases below the threshold rather than above it. I also show that the bunching is more pronounced in trial cases, which suggests that drug amounts are being bumped above the cutoff and not negotiated down to it. Second, I take advantage of the time variation in the crack-cocaine 10-year mandatory minimum threshold to isolate bunching that is solely due to prosecutor manipulation. Finally, I examine data at multiple stages in the criminal justice process and conduct several empirical tests that all suggest prosecutorial discretion negatively affects minority defendants.

[10] In concurrent work, Knorre (2017) finds evidence of bunching in reported drug amounts from Russian police. Specifically, he finds substantial bunching at the minimum amount necessary for prosecution. This result could be driven by police letting offenders off with a warning if they are below a minimum amount. However, Knorre does find a bimodal distribution for seized heroin (though not for seized cannibis or hash) which suggests police may be misreporting those values. Although Knorre does not investigate potential discriminatory behavior or the consequences of the observed bunching, his paper is an interesting complement to this paper in that it finds evidence of the same phenomena in a different setting.

[11] Since Bjerk's paper focuses on sentencing consequences of mandatory minimums for all drug types, he does not empirically investigate the cause of the observed bunching in crack-cocaine offenses. In addition, he does not compare outcomes before and after the Fair Sentencing Act of 2010.

Since I am primarily focused on federal cases, how cases get selected as federal versus state is of critical importance. Whether the crime occurred across state lines is a major determinant in the jurisdiction of the case. If drugs are trafficked across state lines, the case will typically go to federal court. However, cases can be prosecuted federally for a variety of other reasons. Wright (2006) notes that sorting into federal versus state is usually determined by law enforcement agents involved with the case or the prosecuting attorney, but it is never the official purview of judges or defense attorneys. Why might local law enforcement or attorneys wish to pass a case on to the federal courts? For one, local authorities may not have the time or resources to properly pursue a case. Also, Wright suggests that federal sentencing is typically harsher than state sentencing, and that this gap could motivate jurisdiction decisions.

The selection into federal jurisdiction based on the inter-state nature of the crime is unlikely to affect the degree of bunching pre- and post-2010. Local decisions about which cases to pursue and which cases to pass on, however, could affect the degree of bunching. Suppose cases right above the mandatory minimum threshold are particularly difficult and/or time-consuming to win. If this is the case, local authorities may want to leave these cases to the federal courts. However, this explanation suggests the excess mass in cases at or above 280g after 2010 should come solely from cases at or above 50g. In other words, local authorities should be more likely to pass on cases just above 280 grams (thus, the increase in the USSC data after 2010) and less likely to pass on cases just above 50g (thus, a decrease in the USSC data after 2010). The data does show a decrease in cases just above 50g post-2010, but also shows decreases from cases from above 50g. In addition, as long as resources required to win those cases do not differ by race, the observed bunching should not differ by race, yet it does. Finally, the "severity gap" in federal versus state sentencing should not affect bunching at 280g; many state-level thresholds are above 50g, thus cases with 280g or more should be sent to federal jurisdiction even prior to 2010 if local authorities desire harsh punishment. Furthermore, I explore this mechanism empirically in Section III, and I find little support for it as the cause of bunching.

An explanation for increased bunching just above 280g that does comport with the data is that police officers or prosecutors intentionally bump reported amounts above the threshold to increase the probability of a harsher sentence. Law enforcement agents can manipulate drug amounts by choosing the amount of drugs involved in "reverse sting" operations (operations in which agents will sell drugs to an offenders) or by extending traditional sting operations (operations in which agents will buy drugs from offenders) until the total transacted amount is above the threshold (Honold 2014). Outside of these two levers, it is unlikely that law enforcement agents can systematically manipulate drug amounts since evidentiary protocols require the precise logging and controlled storage of evidence.

Likewise, prosecutors can manipulate drug amounts because mandatory minimum sentencing is determined by the amount of drugs the offender is responsible for trafficking, which is not strictly based on the amount of drugs they are holding at the time of arrest (Honold 2014; USSC 2015; Lynch 2016). For one, prosecutors can rely on the testimony of informants or law enforcement to establish “historical weight,” the amount of drugs a defendant is responsible for outside of the actual drugs seized (Lynch 2016). In addition, mandatory minimums also apply to drug trafficking conspiracy crimes in which the total amount trafficked by the group in question can be applied to all members of the group (Sterling 1999). The USSC Federal Sentencing Guidelines (2015) specifically states:

> “Types and quantities of drugs not specified in the count of conviction may be considered in determining the offense level. Where there is no drug seizure or the amount seized does not reflect the scale of the offense, the court shall approximate the quantity of the controlled substance. In making this determination, the court may consider, for example, the price generally obtained for the controlled substance, financial or other records, similar transactions in controlled substances by the defendant, and the size or capability of any laboratory involved.”

The amount of drugs an offender is charged with is at the discretion of law enforcement agents and prosecuting attorneys, thus giving them the opportunity to force the recorded amount just above the threshold. In Section IV, I examine data from local police agencies, the Drug Enforcement Administration, and the Executive Office of the US Attorney prosecutor case management files to locate the source of the bunching. I find evidence that prosecutorial discretion leads to bunching in the case of drug trafficking.

B. Prosecutor Objectives

As discussed above, prosecutors have discretion over the drug quantity charged in federal drug trafficking cases. In addition, the data suggests prosecutors exercise this discretion and that they exercise it differentially by race. In this section, I discuss the literature on prosecutor objectives from the fields of economics, criminology, and law–all of which admit self-interested and/or biased prosecutors. In addition, I introduce a simple conceptual model of prosecutor behavior and charged drug quantities under the assumption that prosecutors aim to maximize sentences.

Since the 1970s, economists have produced several influential theoretical models of plea-bargaining based on prosecutor objective functions. This work began with the canonical economic model of the courts from Landes (1971), which assumes that prosecutors maximize the expected sum of sentences subject to resource constraints. Following Landes (1971), several papers emerged modeling prosecutor objectives as perfectly aligned with society’s objectives. Grossman and Katz (1983), Reiganum (1988), Bjerk (2007), and Baker and Mezzetti (2011) model prosecutors as trying to achieve an ideal punishment for guilty parties and no punishment for innocent parties while facing a resource constraint. Empirical work, however, finds that prosecutors are, in some part, career-focused (Glaeser, Kessler, and Piehl 2000; Boylan 2005). Boylan (2005), for example, shows that for US attorneys longer sentences are associated with positive career outcomes (appointed to a federal judgeship or hired by a large private firm). In addition, recent work demonstrates partisan bias (Nyhan and Rehavi 2017) and racial bias (Rehavi and Starr 2014) in prosecutorial decisions, suggesting that prosecutors may seek harsh punishments for some offenders and lenient punishments for others.

These findings that prosecutors can be self-interested and biased are echoed and often-times preceded by insights from criminologists and legal scholars. Officially, the EOUSA cites Berger v. United States, 295 U.S. 78 (1935) as a description of the role of the US Attorney:

> "The United States Attorney is the representative not of an ordinary party to a controversy, but of a sovereignty whose obligation to govern impartially is as compelling as its obligation to govern at all; and whose interest, therefore, in a criminal prosecution is not that it shall win a case, but that justice shall be done. As such, he is in a peculiar and very definite sense the servant of the law, the twofold aim of which is that guilt shall not escape or innocence suffer. He may prosecute with earnestness and vigor-indeed, he should do so. But, while he may strike hard blows, he is not at liberty to strike foul ones. It is as much his duty to refrain from improper methods calculated to produce a wrongful conviction as it is to use every legitimate means to bring about a just one."

However, the above is more a description of the prosecutorial ideal than the reality. In fact, the case in Berger v. United States, 295 U.S. 78 is itself one of prosecutorial misconduct. Discussions of prosecutorial discretion in law reviews frequently note that career-oriented prosecutors focus on securing lengthy sentences or high conviction rates (Bibas 2004; Simon 2007; Barkow 2009; Sklansky 2017). Since over 90% of federal cases end in plea deals, the plea-bargaining process has also received a great deal of attention in economics, law, and criminology. Stuntz (2004) argues that prosecutors lean on harsh sentences to secure guilty pleas. He even specifically notes the usefulness of sentencing guidelines (e.g. mandatory minimums) in this regard: "plea bargains outside the law's shadow depend on prosecutors' ability to make credible threats of severe post-trial sentences. Sentencing guidelines make it easy to issue those threats." Finally, criminologists and political scientists have also documented prosecutorial bias along race, gender, and partisan lines (Spohn, Gruhl, and Welch 1987; Mustard 2001; Gordon 2009; Shermer and Johnson 2010).

Taking the insights above regarding the objectives of career-focused prosecutors, I build a simple conceptual model to explain how prosecutors decide what drug amount to charge for a given offender. First, assume that for a given defendant d, the prosecutor i observes the physical drug evidence p that was seized by police. The prosecutor maximizes the following utility function:

$$\max_{e} \phi_i(r_d, \omega_d) s^0(a) - c(e)$$

which is increasing in the sentence $s^0(a)$ and is decreasing in the cost of acquiring evidence $c(e)$. The assumption that prosecutors prefer higher sentences is present in Landes (1971) and Yang (2017), is consistent with empirical evidence from Boylan (2005), and reflects legal views on prosecutor decisions in the face of career objectives (Barkow 2009; Sklansky 2017). In addition, prosecutor i's return to a given sentence ϕ_i depends on the defendant's race r_d and a taste shock ω_d. Finally, the sentence $s^0(a)$ depends on the total amount of drugs involved in the case a which is equal to the physical evidence plus the extra evidence e the prosecutor acquires.[12] Assume a simplified sentencing schedule that incorporates mandatory minimums:

$$s^0(a) = \begin{cases} 1 & a < t_L^0 \\ 5 & t_H^0 > a \geq t_L^0, \ a = p + e, \ e \geq 0 \\ 10 & a \geq t_H^0 \end{cases}$$

where t_L^0 is the 5-year mandatory minimum threshold and t_H^0 is the 10-year mandatory minimum threshold. If $p < t_L^0$, then the prosecutor has three sensible options:

1. Choose $e = 0$, leaving the evidence as-is (defendant receives lowest sentence).
2. Choose $e = t_L^0 - p$, adding just enough evidence to reach the 5-year minimum.
3. Choose $e = t_H^0 - p$, adding just enough evidence to reach the 10-year minimum.

Prosecutor i chooses one of the above options to maximize: $\underset{e}{max}\{\phi_i(r_d, \omega_d) \times 1 - c(e = 0), \ \phi(r_d, \omega_d) \times 5 - c(e = t_L^0 - p), \ \phi_i(r_d, \omega_d) \times 10 - c(e = t_H^0 - p)\}$. If $t_H^0 > p \geq t_L^0$, then the prosecutor has two sensible options:

1. Choose $e = 0$, leaving the evidence as-is (defendant receives 5-year minimum sentence).
2. Choose $e = t_H^0 - p$, adding just enough evidence to reach the 10-year minimum.

Prosecutor i chooses one of the above options to maximize: $\underset{e}{max}\{\phi_i(r_d, \omega_d) \times 5 - c(e = 0), \ \phi_i(r_d, \omega_d) \times 10 - c(e = t_H^0 - p)\}$. Finally, the case where $p \geq t_H^0$ is trivial—the prosecutor chooses $e = 0$ since there is no gain to acquiring more evidence. This decision-making process results in excess bunching at t_L^0 and t_H^0.

Now, consider the following change in the sentencing schedule:

$$s^1(p) = \begin{cases} 1 & a < t_L^1 \\ 5 & t_H^1 > a \geq t_L^1, \ a = p + e, \ e \geq 0 \\ 10 & a \geq t_H^1 \end{cases}$$

where $t_L^1 > t_L^0$ and $t_H^1 > t_H^0$ and $t_H^0 > t_L^1$. This will induce the following changes: (1) Some cases previously bunched at t_L^0 will be moved up to t_L^1 (since it is still worth it to bunch at t_L^1) and some will move

[12] Note that e is constrained to be greater than zero. In other words, prosecutors cannot suppress physical drug evidence. Allowing this possibility would result in bunching directly below the mandatory minimum thresholds for defendants that yield a sufficiently negative $\phi_i(r_d, \omega_d)$.

down to $p < t_L^0$ (since it is no longer worth it to acquire extra evidence to bunch at t_L^1); (2) Some cases previously left at $p = [t_L^0, t_L^1)$ will now be worth bunching at t_L^1 while cases previously left at $p = [t_L^1, t_H^0)$ will remain there; (3) Some cases previously bunched at t_H^0 will be moved up to t_H^1 (since it is still worth it to bunch at t_H^1) and some will move down to $p < t_H^0$ (since it is no longer worth it to acquire extra evidence to bunch at t_H^1); (4) Some cases previously left at $p = [t_H^0, t_H^1)$ will now be worth bunching at t_H^1 while cases at t_H^1 or above will remain there.

In Appendix B, I extend the model to include offender responses to the change in the sentencing schedule. The extended model still produces bunching above mandatory minimum thresholds, and disproportionate bunching for black and Hispanic offenders if prosecutors are biased. That said, the data do not show evidence of an offender response to the Fair Sentencing Act. In the Results section, I consider the possibility that the cost of acquiring evidence $c(e)$ differs by race. If the racial disparity in the cost function is specific to each prosecutor, then it is not possible to disentangle prosecutor bias from the prosecutor-specific racial differences in fact-finding costs. However, if all prosecutors (or all prosecutors in a given area) face the same racial difference in the cost of acquiring evidence, then the racial disparity in bunching at 280g should not vary at the prosecutor-level (or at the prosecutor-level within a given area). I find that the racial disparity in bunching varies across states and within Census regions, and that bunching, in general, varies within district.[13]

Absent the extensions discussed above, the simple model predicts that prosecutors will charge an excess number of cases with 280g after 2010, and that those cases would be charged with [50g, 280g) prior to 2010. In addition, the model predicts black and Hispanic offenders will be disproportionately charged with 280g after 2010 if $\phi_i(r_d = [b,h], \omega_d) > \phi_i(r_d = [w], \omega_d)$. Ultimately, this model and the empirical analysis that follows is rooted in broad ideas about prosecutor bias and prosecutors' desire for long sentences, but it also captures a specific phenomenon that has received some attention in law and criminology–federal prosecutors using sentencing guidelines and mandatory minimums to secure guilty pleas or harsh sentences (Stuntz 2004; Honold 2014; Lynch 2016).

In 1983, legal scholar and soon-to-be judge Frank Easterbrook wrote, "Rules could command, for example, that all cases involving a sale of cocaine weighing more than 50 grams be prosecuted and all others not. Rules of this sort produce the arbitrary and unexpected consequences so well known to tax and welfare lawyers; it is far from clear that one can design rules to achieve a particular end. People will change their conduct to take advantage of lacunae." Since then, such rules have been implemented, but researchers have paid scant attention to the ways people have changed their conduct to take advantage of them. In this paper, I document changing conduct by prosecutors that reflects bias against black and Hispanic offenders–behavior that has been discussed and researched qualitatively by legal scholars and criminologists but

[13] Unfortunately, the prosecutor case management files do not include defendant race, so I cannot test for within-district variation in the racial disparity in bunching.

that has remained relatively unexplored empirically.

III. Data

To estimate the degree of bunching at the 10-year mandatory minimum threshold, I use data from the United States Sentencing Commission (USSC) on federal sentencing that includes the amount of drugs involved in the offense. With this data and the change in mandatory minimum thresholds for crack-cocaine in 2010, I can compare the distribution of drug amounts prior to 2010 and post-2010 to estimate the degree of bunching at and above the mandatory minimum threshold. I then bring in several other data sets from different stages in the criminal justice process to determine who is responsible for the bunching at 280g. I describe the USSC data and all additional data used in this paper below.

A. USSC Data

To determine the degree of bunching at or above 280g, I use data provided by the USSC on recorded drug amounts in all federal drug cases sentenced from 1999-2015. These data are summarized in Table 1. I focus on cases that involve a crack-cocaine offense since that is the only drug for which the mandatory minimum changes over time. Approximately 7.8% of offenders in this sample are labeled as white, 10.6% as Hispanic, and 81.6% as black. Table 1 also summarizes information about age, education, citizenship, and details about the offense, all of which are used as covariates in later analyses.

I restrict these data to cases in which the amount of drugs is non-missing and is not recorded as a range. Approximately 20% of cases are excluded for this reason, but the fraction of missing cases for crack-cocaine does not change discontinuously at 2010, though it does increase in 2013 and 2014. Furthermore, in Appendix A, I show that including cases coded as a range only exacerbates the degree of bunching and the racial disparity in bunching (Table A5). I also remove cases that are flagged for having data issues with the drug quantity variable and cases where the court does not accept or changes the findings of fact. Less than 2% of cases are excluded for these reasons.

Using the cleaned data, I plot two histograms (Figures 1a-b) that zoom in on the density around 280 grams for the years before and after 2010. Prior to 2010, the density around 280g is smooth. After 2010, however, 280g becomes the new mandatory minimum threshold and in that same time, the number of cases at and above 280g spikes. Figures 2a-b display how the fraction of cases recorded as 280-290g changes over time. This shows even more clearly that the spike in cases at 280-290g coincides exactly with the policy change. These figures also highlight the racial disparity in bunching at the threshold that occurs after 2010.

B. Additional Data

In addition to data on federal sentences from the United States Sentencing Commission, I incorporate several other datasets to understand the source of the bunching in drug trafficking cases. I describe these datasets here.

Florida and North Carolina State Inmate Databases, 2000-2015.

These data include the year an offender is convicted, a description of the offense, and the offender's race. In Florida, drug offense descriptions typically include the name of the drug involved, and occasionally, the descriptions include a range for the amount of drugs involved. In North Carolina, drug offense descriptions only include the name of the drug involved for trafficking crimes, and in all trafficking cases, the descriptions include a range for the amount of drugs involved. Both states use the range 200-400 grams in classifying offenses. Also, neither Florida nor North Carolina provides information about crack versus non-crack cocaine offenses, describing all such offenses as "cocaine."

National Incident Based Reporting System Property Segment, 2000-2015.

The FBI collects data from local law enforcement agencies about crime, and many agencies report this data at the incident-level. The incident-level reports make up the data in the National Incident Based Reporting System (NIBRS). This data is submitted voluntarily by agencies and thus, it is not representative of national or state-level crime. Upon receipt, the FBI checks the reports for errors and contacts agencies for corrections if necessary. The property segment of this database includes information about drug seizures and drugs involved in arrests.[14]

Drug Enforcement Administration System to Retrieve Information from Drug Evidence (STRIDE), 2000-2015.

The STRIDE database contains information about all drug evidence from DEA and other agencies that was submitted to DEA laboratories for analysis. The data I use was obtained from a Freedom of Information Act request for all records pertaining to the drug "cocaine" from 2000 to 2015. This information includes the year and month the drugs were acquired, the weight of the drugs in grams, and the type of drug (cocaine, cocaine hydrochloride, cocaine base, etc.).

Executive Office of the US Attorney, Caseload Data, 2000-2017.

The Executive Office of the US Attorney (EOUSA) releases case-level data on cases (excluding certain redacted cases) processed by the US Attorney's office. This data is derived from information entered into the Legal Information Office Network System (LIONS). The EOUSA notes that each district may use LIONS differently, and as such, the data should not be used to make cross-district comparisons. Despite these

[14]There are a few well-known issues with NIBRS data. For one, variation in crime definitions over time or across agencies can lead to mismeasurement. This problem, in particular, plagues reporting and measurement of hate crimes and sexual assault (Shively 2005; Bibel 2015). In addition, variation in victim-reporting across agencies can lead to variation in measured crime (Shively 2005). In this paper, I focus on drug crime and drug seizures which are less subject to definitional discrepancies and variation in victim-reporting. Even more, differences in drug seizure reporting across agencies is not problematic in my setting since I am comparing differences in drug amounts pre- and post-2010 and differences across races. There is no evidence of differential reporting or recording of drug seizures pre- and post-2010, in general or by race. Finally, other issues (such as differential coverage and data quality) with NIBRS are covered by Bibel (2015). As far as I can tell, there are no known issues with the drug quantity field of the NIBRS property segment.

restrictions, the data includes a wealth of information about drug cases, including, quantity of the drug and an ID for the lead attorney on the case.

Google Search Trends Data on Racial Animus from Stephens-Davidowitz (2014).

To measure racial animus at the state-level, I use data introduced by Stephens-Davidowitz (2014). Stephens-Davidowitz uses Google search data from 2007 (accessed via the Google Trends tool) and measures relative search volume in every US state for a specific racial slur and its plural form.[15] Since Google searches are virtually anonymous, this measure may provide a less filtered view of racial attitudes than common survey measures. In fact, it is positively correlated with racial animus as measured by implicit association tests or questions about interracial marriage from the General Social Survey. Even more, it is highly predictive of President Obama's vote share in the 2008 and 2012 US elections (Stephens-Davidowitz 2014). The construction of the measure is covered in much greater detail in Stephens-Davidowitz (2014).

IV. Methodology

The approach I use is what Kleven (2016) terms the "difference-in-bunching" method. Many bunching papers, for lack of variation in the threshold of interest, estimate bunching by constructing the counterfactual density from the actual bunched density. To do this, one typically aggregates the data into bins and estimates a regression of the count in each bin on a high-order polynomial of the bin's value and dummy variables for bins in the bunched "window." The estimates from that regression (not including the bunching dummy variables) can be used to predict a smooth distribution of bin counts. Authors then compare that smooth density to the actual density to calculate the degree of bunching in the actual density. In Appendix B, I show that my main results are robust to this method.[16]

Kleven (2016), however, notes that standard bunching estimation is based on a setting where there is no variation in the kink/notch, and calls this a "minimalist approach" that "may not be compelling in all contexts." Additionally, he argues "more sophisticated alternatives exist that require richer data and/or richer variation." The Fair Sentencing Act in 2010 provides this richer variation. In 2010, the threshold for the 10-year mandatory minimum for crack-cocaine changes from 50g to 280g. To quantify the amount of bunching at the new threshold, I compare the post-2010 distribution to the pre-2010 distribution, when

[15] The exact word used is discussed in Stephens-Davidowitz (2014).

[16] To start, I collapse the data on drug quantities for all cases after 2010 to 10 gram bins. I then run a regression of the count of cases on a seventh order polynomial of the bin values and dummy variables for the bins 0-10g, 270-280g, and 280-290g. Then, using the coefficients from the seventh order polynomial and the dummy variable for the bin 0-10g, I calculate a smooth counterfactual distribution. For graphical purposes, I re-scale that smooth distribution to have the same total number of cases as the true distribution. Next, I calculate the percent of all cases that are in the 280-290g bin in the true distribution, the percent of all cases that are in the 280-290g bin in the counterfactual disitribution, and the difference between those two percentages. Finally, I run a regression of the difference between the true and counterfactual distributions on a dummy variable equal to one for the 280-290g bin and equal to zero otherwise (bootstrapped standard errors are calculated by re-sampling the residuals from the polynomial estimation with 200 replications). I carry out a similar procedure to estimate the difference in bunching between white and black and Hispanic offenders (the major difference being that I estimate the counterfactual distributions separately for white and black and Hispanic offenders and that the final regression includes an interaction between the 280-290g bin dummy and a dummy for black and Hispanic offenders).

280g has no special meaning. A number of recent bunching papers take this approach, but despite its rising popularity, there is not a standardized method to estimate bunching in settings with variation in the kink or notch.

Although no standard method exists, most papers using the "difference-in-bunching" approach can be fit into one of two categories. In one, authors estimate bunching using the conventional polynomial method separately for groups where the threshold applies and for groups where the threshold does not apply, using the latter as a placebo test (Best et al. 2015; Fack and Landais 2016; Gelber, Jones, and Sacks 2017; Zaresani 2017; Chen et al. 2018). In the other, authors directly compare the group where the threshold applies to the group where the threshold does not apply. Yet even within the direct comparison category, strategies differ. Several papers compare the distributions by aggregating the data into bins and calculating the difference in levels between the actual and the counterfactual distributions (Brown 2013; Best et al. 2018; Best and Kleven 2018; Cengiz, Dube, Lindner, and Zipperer 2018). Others compare the distributions using regression analysis on the microdata. These papers frequently estimate the difference in the probability an observation is in a given bin between the actual and the counterfactual setting (Kleven et al. 2011; Behagel and Blau 2012; Sallee and Slemrod 2012; Chetty, Friedman, and Saez 2013; Dwenger et al. 2016; Goncalves and Mello 2018; Traxler et al. 2018).

In this paper, I employ both direct comparison methods (aggregate/binned analysis and microdata analysis). I am primarily interested in estimating the change in the probability a case is charged with 280g (or just slightly above 280g) after 2010 and whether that change in probability differs by race. In addition, some analyses in the paper preclude aggregating the data into bins because they rely on data that does not include precise drug quantities. For these reasons, I follow the papers that use regression analysis on microdata to compare the pre- and post-2010 crack-cocaine distributions. For example, I estimate the following linear probability model to calculate the degree of bunching at and above 280g:

$$(Charged\,280-290g)_{it} = \alpha_0 + \beta_1 After2010_{it} + X_i + Y_t + \epsilon_{it} \tag{1}$$

where $(Charged\,280-290g)_{it}$ is equal to one if offender **i** in year **t** is charged with 280-290g (not including 290g) and is equal to zero if the offender is recorded as holding less than 280g or equal to or above 290g.[17] $After2010_{it}$ is equal to one if the offender **i** in year **t** is sentenced in 2011-2014 and is equal to zero if the offender is sentenced in 2000-2010. β_1 then is the change in an offender's probability of being charged with exactly 280-290g as a result of being sentenced after the threshold amount is increased to 280g. X_i represents case-level covariates (such as offender education, race, age, conviction state, etc). and Y_t represent time trends. In most specifications, I limit the sample to 0-1000g to remove extreme outliers and exclude X_i and Y_t, however I show that the result is robust to altering this sample range and robust to

[17] State conviction data does not include precise drug weights. In those cases, I use the dependent variable (Recorded 200-400g).

including numerous controls. To estimate heterogeneity in bunching by race, I use the following model:

$$(Charged\,280\text{—}290g)_{it} = \alpha_0 + \beta_1(After2010 \times White)_{it} + \beta_2(After2010 \times Non\text{—}White)_{it} + Non\text{—}White_{it} + X_i + Y_t + \epsilon_{it} \quad (2)$$

Now, β_1 represents the change in a white offender's probability of being charged with 280-290g as a result of being sentenced after the threshold is increased, and β_2 represents the change for non-white offenders (equal to one if the offender is black or Hispanic).[18] Models (1) and (2) quantify bunching at 280-290g that occurs after the 10-year mandatory minimum threshold changes to 280g

To understand where the excess mass at 280-290g comes from, I use both the binned analysis that compares the counterfactual and actual distributions in levels and regression analysis on the microdata. First, I estimate a series of models similar to the ones above that replace the dependent variable with different drug quantity ranges:

$$(Charged\,X - Y g)_{it} = \alpha_0 + \beta_1 After2010_{it} + X_i + Y_t + \epsilon_{it} \quad (3)$$

In these models, β_1 represents the change in an offender's probability of being charged with an amount of drugs between X and Y grams as a result of being sentenced after the threshold is increased. I estimate model (3) for 0-5g, 5-28g, 28-50g, 50-60g, 60-100g, 100-280g, 200-280g, 290-470g, 470-600g, and 600-1000g. I also estimate model (3) by race. Since the ranges involved are much wider than the 280-290g bin, I include a time trend (centered at zero in 2011) and state fixed effects when estimating equation (3) to partially account for broad differences in drug trafficking over time and across states. Furthermore, to estimate the "jump" in the probability of being within a certain bin after 2010, I estimate the following:

$$(Charged\,X - Y g)_{it} = \alpha_0 + \beta_1 After2010_{it} + \beta_2 After2010 \times Trend_{it} + X_i + Y_t + \epsilon_{it} \quad (4)$$

where β_1 identifies the discontinuity in the time trend at 2011.

To further highlight how the post-2010 distribution differs from the pre-2010 distribution, I aggregate the cases into 10g bins pre- and post-2010. Following Best et al. (2018), I estimate 90% confidence intervals with a bootstrap procedure that samples cases with replacement from the microdata.[19] I compare the binned distributions to estimate the net change in bins below 280g, at 280-290g, and above 290g. This analysis, in addition to the microdata analysis, addresses a critical question for policy implications: how would offenders who were charged with 280-290g post-2010 have been charged pre-2010? If those

[18] I typically refer to "black and Hispanic offenders" directly, but for brevity in the tables and equations, I use the term "non-white."

[19] I draw 50 random samples from the microdata and do the binned analysis on each sample. The final number of cases for each bin is calculated as the mean of the number of cases across all 50 samples, and the final standard error is calculated as the mean of the standard error across all 50 samples.

offenders would have been charged below 280g, then the bunching at 280-290g post-2010 may represent an effort to increase sentence lengths for some offenders.

In the Results section, I detail methodology and results for several additional analyses. Section V.B discusses the sentencing consequences of bunching. In Section V.C, I investigate three potential mechanisms that could explain the observed bunching. Section V.D examines bunching from prosecutors before and after a Supreme Court decision that changed the evidentiary standard for mandatory minimum sentencing. Finally, Section IV.E tests alternative explanations for the main results and for heterogeneity in bunching by offender characteristics and state-level racial animus.

V. Results

A. Main Results

1. Primary Bunching Estimates and Robustness

Using sentencing data from the USSC, I estimate the effect of being sentenced after 2010 on whether an offender is sentenced for a drug amount between 280-290g. Column 1 of Table 2 indicates that offenders sentenced after the threshold increases are more likely to be charged with amounts just above the threshold. An offender sentenced after 2010 is 3.5 percentage points more likely to be charged with a drug amount between 280-290g. Column 2 shows that this increase in bunching is driven by black and Hispanic offenders, and column 3 shows that this result is robust to removing Hispanic offenders from the sample.

Furthermore, this result is robust to various sample restrictions, the inclusion of state and time controls, the inclusion of offender-level controls (criminal history, age, etc.), clustering standard errors at the state-level, the use of Logit/Probit models instead of linear probability, wider bunching ranges (280-320g, for example), and the inclusion of narrow bandwidth dummy variables (see Tables A1-A5). Finally, Figure A1 shows this effect, in general and by race, for each year relative to 2010 and Figure A2 provides further visual evidence by displaying the ratio of the percent of cases in each 10-gram bin post-2010 to the percent of cases in each 10-gram bin pre-2010 (if the distributions were the same, this ratio would equal 1).

Aggregate bunching analyses also yield similar results. Figure 3a plots the counterfactual (scaled pre-2010) density and the actual post-2010 density. The spike at 280g in the post-2010 density is the bunching that is detected in Table 2. The results from equation (4) indicate that after 2010, there is a 3.5 percentage point increase in cases with 280-290g ($\beta = 0.0349$ and SE $= 0.0021$). Figures 3b-c show the densities by race.[20] The bunching at 280g in the post-2010 density is larger for black and Hispanic offenders. In fact, the results from equation (5) indicate that after 2010, the rise in cases with 280-290g is about 2 percentage points higher for black and Hispanic offenders ($\delta = 0.0197$ and SE $= 0.0094$).

[20] To make these figures easier to read, I limit the cases to 0-500g. I find similar results with 0-1000g.

The binned plots in Figures 3a-3c reveal a spike in the number of cases at 50-60g both before and after 2010. Prior to 2010, the 10-year mandatory minimum threshold for crack-cocaine is at 50g, and thus, excess mass at and above the threshold before 2010 could be the result of prosecutorial discretion. After 2010, however, there is no threshold at 50g. In this case, the persistent excess mass at 50g is likely due to round-number bias from offenders, law enforcement, or prosecutors. The powder cocaine distribution, which never has a mandatory minimum threshold at 50g, exhibits similar excess mass at 50g. For crack-cocaine, the fraction of cases from 50-60g is about 1.5 times the fraction of cases from 40-50g (not including 50g). For powder cocaine, that ratio is similar–the fraction of cases from 50-60g is about 1.7 times the fraction of cases from 40-50g. While conventional bunching estimation techniques would address the presence of round-number bias by accounting for it in the estimation of the smooth polynomial fit, the difference-in-bunching method accomodates round-number bunching directly because that bunching will be present in both the counterfactual and actual distributions (Best et al. 2018).

2. Source of the Excess Mass at 280g

To understand the reason for this bunching, I analyze other parts of the drug quantity distribution. If the excess mass in 280-290g after 2010 comes from above 290g, this bunching may be the result of negotiation between prosecutors and defendants (Bjerk 2016). However, if the excess mass comes from below 280g, it is possible that prosecutors are shading reported amounts upwards to exceed the threshold amount. Figure 4 plots the difference between the post-2010 and the scaled pre-2010 densities for each 10g bin and adds confidence intervals by using 200 bootstrapped samples from the microdata.[21] When this difference is below zero, it means the bin contains relatively fewer cases after 2010 and when the difference is above zero, it means the bin contains more cases after 2010.

The figure shows an increase of about 340 cases in the 280-290g bin post-2010, a net increase in cases above 280g, and a net decrease below 280g. Summing the changes in bins above 280g, I find a net increase in that section of the distribution after 2010. The point estimate on the net change is noisy, but even summing the lower bound of the confidence interval for all bins above 280g can only account for about 46% of the increase in the 280-290g bin. On the other hand, the net change below 280g can account for 120% of the increase in the 280-290g bin. Again, this point estimate is noisy. In fact, summing the upper confidence interval for all bins below 280g implies a net increase in that section of the distribution. The key takeaway is that changes in the distribution below 280g can account for the excess mass at 280g, whereas changes in the distribution above 280g cannot. In other words, an offender charged with 280-290g

[21] In Appendix A, I show similar results using a linear probability model with the microdata. The probability an offender is charged with a drug weight below 280g decreases after 2010, whereas the probability an offender is charged with a drug weight above 280g increases after 2010. In general, it is hard to distinguish these changes from long-run trends since they are not as sharp as the increase in 280-290g. To help with this, I plot the share of cases in several different bins below 280g and above 290g.

post-2010 would likely have been charged with less than 280g had they been sentenced prior to 2010. [22]

One concern with this figure is that most of the missing mass below 280g appears to come from below 50g. First, even the changes from 50g-280g can account for 85% of the increase in the 280-290g bin. Second, it is possible that offenders charged with 280-290g post-2010 would have been charged with below 50g pre-2010. The simple conceptual model in Section II.B does not allow for this possibility, but introducing a fixed cost of evidence-gathering could explain this behavior. For example, if an offender is caught with 10g of physical evidence prior to 2010, it may not be worthwhile to collect evidence to push them from a 5-year sentence to a 10-year sentence. After 2010, however, that same offender would face a 1-year sentence without some additional evidence-gathering. Once prosecutors pay the fixed cost to gather evidence, it may now be worthwhile to gather enough to get the 10-year sentence. In other words, it is entirely possible that decreases from 0-50g contribute to the increase in cases from 280-290g.[23]

In Figures 5a-i, I plot the share of cases over time in each of the following ranges: 0-5g, 5-28g, 28-50g, 50-60g, 60-100g, 100-280g, 290-470g, 470-600g, and 600-1000g. Figures 5b, 5d, 5e, and 5f, in particular, show a break in the share of drugs from 5-28g, 50-60g, 60-100g, and 100-280g after 2010. Figures 5j and 5k show the total share of cases below 280g over time and the total share above 280g over time. For all of these shares, there are considerable trends over time. To quantify the break in the trend at 2011, I estimate case-level regressions that interact the dummy variable for after 2010 with a linear time trend centered at zero in 2011. These results are in Table 3a.[24] In Table 3b, I estimate case-level regressions for the probability of being below 280g or above 290g using various trend interactions. All of these regressions also include state fixed effects to account for regional differences. The probability of being sentenced for amounts below 280g decreases after 2010 while the probability of being sentenced for amounts above 290g increases. In addition, column (2) in Panel A of Table 3a indicates that offenders are more likely to be charged with 28-50g after 2010. This may reflect bunching due to the changing 5-year mandatory minimum threshold from 5g to 28g.

Finally, I examine the degree of bunching in the subset of cases that go to trial. If the bunching is a result of lenient prosecutors rounding down, we should expect less bunching in trial cases where incentives for leniency are muted. However, the degree of bunching and the racial disparity in bunching is only heightened in trial cases. In fact, the only cases with 280-290g that go to trial are those of black and

[22] In Appendix A, I show that this result holds for black and Hispanic offenders alone (the below/above differences are not driven by white offenders) and that the result is robust to netting out the median annual change in each 10g bin pre-2010 (the changes in those bins are not typical yearly changes). I also show a plot of the differences after netting out the 25th percentile of annual changes in each 10g bin pre-2010. In this figure, I still find more missing mass below the threshold than above the threshold. However, the 50-280g range no longer sums to a net decrease on its own.

[23] In fact, anecdotal evidence suggests prosecutors do push offenders from low levels of physical evidence to high charged amounts. In *Hard Bargains*, Mona Lynch recounts a quote from an anonymous AUSA, "The actual heroin sales directly tied to Mr. Samuels and his son were of 1 gram and 4 grams, respectively; the rest was arrived at on the mere say-so of confidential informants. [...] She (the assistant U.S. attorney prosecuting the case) told me that she could have established enough historical weight, through those (conspirators) she had 'flipped,' to get Mr. Samuels to at least a ten-year mandatory minimum sentence, if not more."

[24] Results without the trend interaction and results with a quadratic trend interaction are in Appendix Tables A6-A7.

Hispanic offenders. Column 4 of Table 2 indicates that bunching also increases in trial cases post-2010, and as before, the increased bunching is accompanied by a falling share of cases below 280g and a rising share of cases above 290g. This is further evidence that the observed bunching is a result of shading up rather than negotiating down.

Ultimately, these results show an increase in the fraction of cases at and above 280g after the 10-year mandatory minimum threshold shifts to 280g in 2010. This increase is almost three times larger for black and Hispanic offenders than white offenders. Comparing the pre- and post-2010 distributions of drug amounts used in Federal sentencing suggests the increase in cases at and above 280g is due to cases that would have been charged below 280g prior to 2010. In Section V.C, I evaluate three potential explanations for these patterns in the USSC sentencing data, and I find that a small fraction of prosecutors are using their discretion to tag offenders with this amount that triggers the 10-year mandatory minimum. First, I check that this bunching actually matters for sentencing. In other words, do offenders charged at 280-290g after 2010 receive longer sentences than offenders charged below 280g after 2010?

B. Sentencing Consequences

In order to understand the policy implications of this bunching, I estimate the sentencing consequences of crossing the mandatory minimum threshold. Since mandatory minimum sentencing only gives guidelines about minimum sentencing, it is possible that being above the amount has no affect on actual sentencing. Judges could treat defendants with 270g the same as defendants with 280g and apply the mandatory minimum sentence of 10 years to both. In that case, the existence of bunching might be trivial since no extra penalty is applied to those offenders who are "bunched" at 280g. However, it is also possible that crossing the mandatory minimum threshold does lead to harsher sentences because judges do not have as much room for leniency once a case triggers the mandatory minimum. I investigate these possibilities with an regression discontinuity style regression:

$$\begin{aligned} Sentence_{it} = \alpha &+ \beta_1 Above280_{it} + \beta_2 Amount_{it} + \beta_3 (Above280 \times Amount)_{it} \\ &+ \delta_1 (Above280 \times After2010)_{it} + \delta_2 (Amount \times After2010)_{it} \\ &+ \delta_3 (Above280 \times Amount \times After2010)_{it} + Y_t + \epsilon_{it} \end{aligned} \tag{5}$$

where $Sentence_{it}$ is the sentence handed down for offender **i** at time **t**, $Above280_{it}$ is equal to one if the defendant is recorded with 280g or more of crack-cocaine and zero otherwise, $Amount_{it}$ is equal to the defendant's recorded drug quantity centered at 280g, $After2010_{it}$ is equal to one if the sentence occurs after 2010 and zero otherwise, and Y_t is a linear time trend. The coefficient δ_1 identifies the sentencing penalty from crossing the 280g mandatory minimum threshold since that threshold is only in effect after

2010. I also estimate similar regressions including the 50g threshold that is in effect prior to 2010. As long as the offenders who are bunched above the threshold are not negatively selected from the population just below the threshold, this methodology will provide a conservative estimate of the sentencing penalty.[25]

I find that crossing the 280g threshold does have sentencing consequences. Offenders recorded with 270-280g after 2010 have a mean sentence of 9.6 years whereas offenders recorded with 280-290g after 2010 have a mean sentence of 11.2 years. Table 4 provides formal estimates of the effect of being charged with at or above 280g. Columns 1 and 2 indicate that after 2010, offenders with at or above 280g of crack-cocaine had about 1.75-2.15 years added to their sentence. Likewise, after 2010, offenders with at or above 50g of crack-cocaine faced sentences about 1.19-1.75 years lower than those offenders prior to 2010.[26] Column 5 estimates a model including both thresholds and finds similar results. Finally, Column 6 includes offenders who received life sentences (coded as 70 years) and offenders who received sentences less than one month, and again, I find that offenders recorded with 280g or more face harsher sentences.

It is possible that the offenders in the bunching range are negatively selected from offenders below 280g, or in other words, are offenders who would receive harsher sentences even in the absence of bunching. I test this possibility by comparing the criminal history scores of offenders in the bunching range to those offenders below the bunching range. In fact, I find that offenders with 280-290g after 2010 have lower criminal history scores than offenders just below 280g. This suggests that, if anything, those offenders who are bunched may be positively selected from the distribution below 280g. This is likely due to sentencing incentives–offenders with higher criminal history scores face harsher sentences regardless of the quantity of drugs they are reportedly involved with. Ultimately, this further suggests that the estimated sentencing penalty from equation (6) is a conservative estimate of the sentencing consequence of bunching.

In Figure 6a, I plot sentencing outcomes by drug weight from 230-330g and the linear fit on each side of the 280g threshold for cases sentenced after 2010. This provides visual evidence of the sentencing penalty from crossing the mandatory minimum threshold. Figure 6b shows this same plot for the subset of cases sentenced in states that have low levels of bunching. Even in these states, where there is little manipulation around the threshold, there is a sentencing penalty of about 1.8 years.[27]Again, this estimate assumes that an offender bunched at 280g would be charged with an amount just below 280g in the absence of the 280g

[25]The estimate is conservative because it assumes offenders who are bunched at 280-290g would be recorded just below 280g in the absence of bunching. If offenders are bumped above 280g from throughout the distribution, then the sentencing consequence would be harsher.

[26]I do not find significant differences in these sentencing discontinuities by race.

[27]Recent papers by Diamond and Persson (2017) and Dee et al. (2017) study the consequences of test score manipulation. Both papers address the issue of selection into bunching. Dee et al. take advantage of their unique setting. Due to changes in how tests were scored, manipulation was easier in some years than others. In years when scoring was centralized and no re-scoring was allowed, there is little to no manipulation. Dee et al. exploit this time variation in the prevalence of bunching to estimate the effect of bunching on outcomes. Unfortunately, my setting does not allow for this approach since manipulation of reported amounts is always possible (even pre-2010 when the threshold is 50g instead of 280g). Diamond and Persson take a different approach. To deal with selection, they use test scores outside an estimated "manipulable range" to construct a counterfactual outcome distribution. The authors then compare the actual outcome distribution to the counterfactual outcome distribution to determine the impact of manipulation. Again, this approach isn't suitable for my setting since the entire range of amounts is manipulable.

threshold. However, the results above imply that offenders bunched at 280g come from throughout the distribution below 280g. The average sentence after 2010 for offenders in the 50-280g range is 7.9 years, and the average sentence after 2010 for offenders in the 0-280g range is 7.3 years. Using those values for the counterfactual sentence implies a sentencing consequence of 3.3 years and 3.9 years respectively.

C. Potential Mechanisms

The three mechanisms I evaluate are: (1) a shifting composition of cases between state and federal court, (2) law enforcement discretion, and (3) prosecutorial discretion. For these analyses, I rely mostly on visual evidence, but a formal analysis showing the main bunching results for each mechanism is in Table 5.

First, I show the fraction of cocaine convictions in Florida recorded as 200-400 grams over time. If the bunching in federal cases is due to state and local authorities sending more 280 gram cases to federal prosecutors, then there should be a decrease in the fraction of cases with 200-400g after 2010.[28] Even more, this decrease should be especially pronounced for black and Hispanic offenders. I do not find a decrease in state convictions for 200-400g in general or by race. This implies shifting cases from state to Federal court cannot explain the bunching at 280g. I provide similar visual evidence to evaluate mechanisms (2) and (3). If police are the source of the bunching in federal cases, then this bunching should be evident in law enforcement data on seized drug quantities. I do not find evidence of bunching from drug seizure records. Finally, if prosecutors are the source of the bunching, then I should observe the bunching only in case management data from the Executive Office of the US Attorney (EOUSA). I do, in fact, find bunching in prosecutor case management files. The presence of bunching at the prosecutor level but not at an earlier stage implies that prosecutors are responsible for the excess mass at 280g after 2010.

1. Shifting of Cases Between State and Federal Courts

Drug Convictions in Florida and North Carolina Courts

The data used in the preceding analyses come from federal sentencing, and it is possible that the types of cases prosecuted in federal court changes after 2010. In order for this shift to affect bunching, it must be the case that state and local authorities send more cases just above 280g to federal courts after the threshold changes in 2010. To test this, I use state-level data on cocaine offense convictions from Florida.[29,30] Specifically, I test for a shifting composition in the type of cocaine offenses over time by using offense descriptions from Florida which specify the range of the drug quantity involved in the offense. Unfortunately,

[28]This is true as long as the increase in 280-290 gram cases is not accompanied by a sharp decrease in cases from 200-280 grams or 290-400 grams. I show visual evidence that this is not the case. Even in the broad weight range of 200-400 grams, there is an increase post-2010 in federal cases.

[29]In Appendix A, I show similar results for North Carolina. I do not include NC in the main analysis because many of its drug convictions do not include any information about drug type involved.

[30]Unfortunately, these data do not distinguish between crack and non-crack cocaine offenses. Data from Missouri indicates that approximately 80% of all cocaine offenses are crack-cocaine offenses.

Florida classifies drug offenses using broad ranges: 0-28g, 28-200g, 200-400g, and 400+g. However, the USSC data show a sharp 3.5 percentage point increase in cases with 200-400g after 2010 (plotted in Figure 7c). Thus, we should still see substantial decreases in the broad 200-400g after 2010 if shifting from state to federal court is the cause of the observed bunching.

Figures 7a-7b plots the share of all cocaine cases in Florida that are for offenses with 200-400g of cocaine (for all offenders and by race). I do not find that the composition of cases changes after 2010 in general or by race.[31] Columns 1 and 2 of Table 6 confirm this. The probability a state-level drug conviction is in the 200-400g range in Florida does not meaningfully change after 2010. In fact, for black and hispanic offenders, there is an increase in the fraction of cases with 200-400g. This suggests that a shifting of cases between state and federal court does not explain the bunching in drug trafficking.

Since there are many more cases convicted at the state-level versus Federal-level, it is possible that a minor, undetectable shift in Florida would be detectable at the Federal-level. This is not the case for the 200-400g range. First, the state-Federal disparity in number of cases is due to states prosecuting more minor possession cases than the Federal courts. There are 150 crack or powder cocaine cases in the 200-400g range convicted in Federal court districts located in Florida after 2010. There are only 200 cases in this range convicted in Florida state courts after 2010. Re-coding 150 of the 200 Florida cases as not in the 200-400g range does yield a detectable effect. Similarly, recoding 150 cases not in the 200-400g range as in the 200-400g range also yields a detectable effect.[32]This simple simulation implies that a shift of cases from Florida to the Federal system would be detectable.

Bunching by Law Enforcement Agency Sending Case to EOUSA

The EOUSA prosecutor case management files (which are analyzed in more detail below) include a field that indicates the law enforcement agency that sends the case. If the bunching at 280g is caused by a shift from state courts to federal courts, then bunching should only be present in cases with state law enforcement involved. In Figure A7, I plot the fraction of cases with 280-290g over time by the type of agency involved. I find that bunching at 280g is present in cases with state law enforcement involvement but also in cases that are sent from Federal agencies. This is further evidence that the bunching at 280g after 2010 is not the result of state to federal case shifting.

2. Law Enforcement Discretion

NIBRS, Local Law Enforcement Drug Seizures

Local police departments voluntarily report incident-level information to the FBI about drug seizures or

[31]Figure AX plots the 200-400g cases as a share of all cocaine offenses in Florida that specify a weight. Results are similar.

[32]A related concern is that the large number of cases in urban counties may mask shifting in rural counties. I split the analysis by counties with greater than 5000 cocaine convictions from 2000-2015 and counties with less than 5000 cocaine convictions from 2000-2015. I do not find substantial shifting for either group. For small counties (those with less than 5000 cocaine convictions), I find a decrease in cases with 200-400g of about 0.1 percentage points. For large counties (those with more than 5000 cocaine convictions), I find no change in cases with 200-400g (less than 0.02 percentage points).

the quantity of drugs involved in arrests. This data is then reported for public use in the National Incident Based Reporting System Property Segment. Using the NIBRS data on drug crime, I create a balanced panel of agencies from 2000-2015 and examine the distribution of drug seizure quantities.[33] If local law enforcement is the source of bunching, I should observe an increase in bunching at 280-290g after 2010. Figure 8 plots the fraction of drug seizures with 280-290g over time and does not show an increase in drug seizures with 280-290g after 2010, in general or by race. These results are also shown in Columns 3 and 4 of Table 5.[34] Finally, only 5 incidents total are reported in the NIBRS after 2010. This suggests that discretion in local law enforcement and drug "sting" tactics cannot explain the bunching in drug amounts after 2010.

DEA STRIDE, Federal Law Enforcement Drug Seizures

I also test for bunching in drug quantities from the Drug Enforcement Administration's System to Retrieve Information from Drug Evidence (STRIDE) database.[35] This data includes exhibits sent to DEA laboratories from both federal and local law enforcement agencies. Figure 9 plots the share of cocaine exhibits with weights from 280-290g from 2000-2015. There is no increase in exhibits with 280-290g after 2010. Again, Table 5 also shows this result. In fact, there are less than 20 total cocaine exhibits in the DEA data with 280-290g after 2010. This further suggests that local and federal law enforcement are not responsible for the observed bunching at 280g after 2010.

3. Prosecutorial Discretion

Bunching in Prosecutor Case Management Files

The Executive Office of the US Attorney provides case-level data extracted from an internal case management system. Using this data, I test for bunching in the quantity of drugs recorded in the case management system. Figure 10 shows that there is a sharp increase in the fraction of cases recorded with 280-290g after 2010. Since I find no evidence of bunching in data from law enforcement, this suggests that the bunching occurs once the case is in the hands of the prosecutor.

Table 5 indicates that the fraction of cases in 280-290g increases by 7.7 percentage points after 2010. This is twice the increase I find in the sentencing files. This difference is likely driven by missing values in the EOUSA files. Re-coding each missing value as though it were not in the 280-290g range yields an increase of about 3.5 percentage points after 2010, which is more in line with estimates from the sentencing

[33] I also estimate the degree of bunching using only states that have full coverage (i.e. states in which all agencies are participating in NIBRS). I still observe bunching in final federal sentencing for cases convicted in these states, but again, I do not find any evidence of bunching in drug seizures for these states. This mitigates concerns that bunching in the NIBRS data is masked by differential participation by agencies over time.

[34] These results are based on a balanced panel of agencies from 2000-2015. If agencies that would bunch at 280g drop out of the sample after 2010 or only join the sample after 2010, then this analysis will fail to uncover law enforcement bunching. In Appendix A, I show results based on states where all agencies participate in NIBRS. I do not find bunching at 280g in these "full coverage" states either.

[35] The analysis in this section uses unvalidated DEA data, and I claim authorship and responsibility for all inferences and conclusions that I draw from this information.

data. In general, I ignore this missing value problem and use the data as it is recorded. However, for cross-district comparisons, it is important to incorporate the re-coding since the use of missing values varies by district. In Appendix A, I show the main results below are robust to the missing value re-coding.

Prosecutor-level Bunching Estimates

To explore bunching by prosecutors further, I use the ID of the lead attorney on each case and test for heterogeneity in bunching by attorney. Since each attorney only has a small number of cases and since I do not know the specific circumstances of each case, I cannot pinpoint "bad behavior" from any individual attorney. However, by estimating bunching separately for each attorney, I can calculate the fraction of prosecutors responsible for the observed bunching. Also, I can compare the distribution of cases for bunching and non-bunching attorneys to further understand where the excess mass at 280-290g is coming from.

Prior to 2010, approximately 0.4% of all cases with a drug quantity less than 1000g were recorded as having 280-290g. I use this statistic as a benchmark to detect attorneys who bunch after 2010.[36] For each attorney, I calculate the percentage of their cases with 280-290g of drugs after 2010. Then, I define bunching for each attorney as the difference between that percentage and 0.4%. Specifically, the "bunching metric" for attorney i is:

$$\text{Bunching Metric}_i = \text{\% of Cases with 280-290g}_i\text{-0.4\%} \tag{6}$$

Figure 11 plots a histogram of the resulting measure for the 94 attorneys who served as lead attorney on at least 15 drug cases after 2010.[37] The majority of these attorneys (about 70%) exhibit little to no bunching. In other words, their bunching metric is equal to or above -0.4% but below 0%, indicating that their fraction of cases with 280-290g was at or below the pre-2010 average. Approximately 30% of prosecutors, however, do have a higher than normal percentage of cases with 280-290g after 2010. While Figure 11 only plots the bunching metric for attorneys with 15 or more drug cases post-2010, this ratio of non-bunching to bunching attorneys holds for the whole data. In fact, the bunching attorneys represented in the Figure 11 (those with 15+ cases post-2010) do not even account for half of the total observed bunching. For example, removing all bunching attorneys with 15 or more cases post-2010 only lowers the bunching coefficient from 0.08 to 0.06. A large part of the observed bunching is accounted for by prosecutors with fewer than 15 drug cases after 2010. In Figure 14c, I map the number of bunching attorneys in each state (among attorneys with 5 or more drug cases post-2010).

This individual-level bunching is also persistent over time. There are 21 attorneys who bunch at 280-

[36]In Appendix A, I use the district-level pre-2010 average to account for district fixed effects in cases at 280-290g. I can also use each attorney's pre-2010 behavior as their own benchmark to detect bunching post-2010. This approach yields similar results, but it limits the number of attorneys I can classify since very few attorneys have a sufficient number of drug cases both pre- and post-2010.

[37]Results are similar when using lead attorneys with 5 or more drug cases after 2010.

290g post-2010 and who serve as lead attorney on 15 cases prior to 2010 and on 15 cases after 2010. Of these 21 attorneys who bunch at 280-290g post-2010, 20 of them also exhibit bunching at 50-60g pre-2010. Furthermore, the attorney-level bunching cannot be accounted for by district fixed effects. The average within-district standard deviation in the 280-290g bunching metric is 0.13, and district fixed effects only explain about 6% of the variance in the bunching metric. This suggests that the heterogeneity in bunching at the attorney-level is not due to district-level differences alone.

Further Evidence on Source of Excess Mass at 280g

Finally, in Figure 12a, I plot the post-2010 density of drug weight for the non-bunching attorneys and the post-2010 density for bunching attorneys. This echoes the approach that Goncalves and Mello (2018) use to formally estimate bunching in ticketed speeds.[38] The non-bunching attorneys provide an alternative counterfactual density since they are not responding to the mandatory minimum thresholds in the same way as the bunching attorneys. Comparing these two densities, I see that non-bunching attorneys have an excess mass of cases below 280g. This is particularly evident in Figure 12b where I plot the difference between the two densities from Figure 12a. Below 280g, the differences are almost entirely negative. In other words, the attorneys who bunch at 280g have relatively fewer cases below 280g than those attorneys who do not bunch at 280g. I do not see substantial differences in the densities above 280g. This provides further evidence, from different data and a different source of variation, that those attorneys who bunch are shading up the reported quantity of crack-cocaine.

While it may be surprising that prosecutors could induce this bunching, this ability is explicitly written into federal sentencing guidelines. The total drug quantity used to determine sentencing is not strictly tied to the amount found on the defendant at the time of arrest. Instead, the court considers all quantities relevant to the count of conviction. Attorney Dan Honold, writing in the Harvard Journal on Legislation, states, "the rule that a defendant is responsible for the entire quantity of the controlled substance found relevant to the conviction, leads to higher sentences. This is so because in cases where a defendant is found to be involved with a higher quantity of a controlled substance than he or she physically possessed at the time of arrest, that quantity will be added to the quantity calculation under Subsection (a)(5)." This rule gives prosecutors discretion to build a case about "relevant" quantities, and present it to the presiding judge. Even more, prior to June 2013, the evidence about relevant quantities did not need to satisfy the "beyond a reasonable doubt" evidentiary standard, because the "principles and limits of sentencing accountability under this guideline are not always the same as the principles and limits of criminal liability (USSC, 2015)." A Supreme Court decision in June 2013 changed the evidentiary standard, and as a result, partially reigned in the observed bunching, but it did not eliminate it entirely.[39]

[38] They compare lenient police officers to non-lenient police officers.

[39] Mona Lynch, a criminologist at UC Irvine, has compiled qualitative evidence about the reach of federal sentencing guidelines in her book *Hard Bargains*. Lynch finds that prosecutors use informants to establish "relevant" quantities: "Informants provide information that is used to estimate drug weight for alleged past trafficking acts—they tell the case law enforcement agents how much

D. The Impact of *Alleyne v. United States*

On January 14, 2013, the Supreme Court began hearing arguments in the case Alleyne v. United States. The petitioner, Allen Alleyne, argued that facts that increase the mandatory minimum sentence for a defendant are "elements" of the alleged crime and should be evaluated by a jury. In a 5-4 decision on June 17, 2013, the Court ruled in favor of Alleyne and issued a decision that changed the evidentiary standard for evidence related to mandatory minimum sentencing enhancements (Bala, 2015).

Prior to this decision, evidence on drug quantities was presented to the judge during the "sentencing phase" of a trial. The presiding judge would then decide, based on the legal standard of "a preponderance of evidence," whether the mandatory minimum sentence applied. The Supreme Court decision required that evidence that would raise the minimum sentence for an offender be presented to the jury and evaluated based on the stricter legal standard of "beyond a reasonable doubt." I estimate how prosecutors reacted to this decision by comparing the change in bunching around June 17, 2013 to the change around June 17th in other years after 2010. If prosecutors are inflating drug amounts to levels that could not be supported at trial, then there will be a decrease in bunching for cases received after the Supreme Court decision.

Using the EOUSA case management data, I implement a difference-in-discontinuities design that compares the discontinuity in the prevalence of bunching for cases received around June 17, 2013 to the discontinuities for cases received around June 17 in all years after 2010 excluding 2013. I estimate the following:

$$\begin{aligned}(Recorded\,280-290g)_{it} &= \alpha_0 + \beta_1 AfterJune17_{it} + \beta_2 DaysFromJune17_{it} + \beta_3(After \times DaysFrom)_{it}\\ &+ \delta_1(AfterJune17 \times Year2013)_{it} + \delta_2(DaysFromJune17 \times Year2013)_{it}\\ &+ \delta_3(After \times DaysFrom \times Year2013)_{it} + D_{it} + \epsilon_{it}\end{aligned}$$

where $After_{it}$ is equal to one if case **i** is received after June 17th of year **t** but before January 1st of year **t+1** and is equal to zero if case **i** is received before June 17th of year **t** but after January 1st of year **t**. $DaysFrom_{it}$ is the number of days from June 17th that case **i** is received, and $Year2013_{it}$ is equal to one if case **i** is received in 2013 and is equal to zero if it is received in 2011-2012 or 2014-2016.[40] D_{it} represents day-of-week fixed effects. The coefficient β_1 is the average discontinuity in the fraction of cases with 280-290g after June 17 from 2011-2016. The coefficient δ_1 is the discontinuity that is specific to June

was sold, how often, and for how long. This information constitutes historical weight—no drugs have to be found or tested or put on the scale for it to be the basis of conspiracy convictions and subsequent "relevant conduct" at sentencing. And it can dramatically increase sentencing exposure against those informed upon." Lynch even writes about a situation in which an assistant US attorney directly describes how relevant quantities can be established: "The actual heroin sales directly tied to Mr. Samuels and his son were of 1 gram and 4 grams, respectively; the rest was arrived at on the mere say-so of confidential informants. [...] She (the assistant U.S. attorney prosecuting the case) told me that she could have established enough historical weight, through those (conspirators) she had 'flipped,' to get Mr. Samuels to at least a ten-year mandatory minimum sentence, if not more. So on top of the 100 grams of heroin alleged in the conspiracy charge, she indicated she could also have tagged Samuels and his son with 5 kilograms of cocaine."

[40] I do not include 2017 in these analyses since the data do not include the full year.

17, 2013–the date of the *Alleyne* decision.[41,42]

Column 2 of Table 6 shows this result using a bandwidth of 130 days (the Imbens-Kalyanaraman optimal bandwidth) before and after June 17th in each year. The coefficient in the first row indicates that, on average, there is approximately no change in bunching after each June 17th from 2011-2016. The next coefficient, labeled "After June 17, 2013", shows the change in bunching that is specific to June 17, 2013. I find that bunching changes discontinuously only after June 17, 2013. In fact, the fraction of cases recorded with 280-290g drops by about 15 percentage points after the ruling in Alleyne. This is also the case for the 120-day and 60-day bandwidth, although as I narrow the bandwidth, I lose precision.[43]

Figure 13 illustrates why there is a large discontinuity in the fraction of cases with 280-290g around June 17, 2013. In the run up to the Supreme Court's decision, the fraction of cases that are bunched right above 280g balloons to levels unseen in other years. It is not clear if this increase is due to a natural rise in the practice of bunching or if the increase is due to prosecutors rushing to exercise their discretion before it was limited by the Court. Regardless, it is clear that Alleyne at least somewhat reigned in the practice of bunching. This suggests that prosecutors were using discretion to build cases on evidence that was unlikely to pass "beyond a reasonable doubt" scrutiny from juries.

E. Alternative Explanations and Heterogeneity in Bunching

1. Offender Behavior

In Appendix C, I consider a conceptual model of offender behavior in response to the Fair Sentencing Act in 2010. If black and Hispanic offenders respond differently than white offenders to the Fair Sentencing Act, a racial disparity in bunching at 280g may reflect prosecutors' reactions to those different responses rather than racial discrimination. In Tables 7a-7b, I show that, in general, white, black, and Hispanic offenders are arrested with similar drug quantities and are similarly likely to be holding 280-290g. In addition, I show that black and Hispanic offenders are not arrested with more drugs following the Fair Sentencing Act, but instead, are holding smaller amounts when arrested after 2010. This implies that the racial disparity in bunching cannot be attributed to differential responses by race.

[41] In response to *Alleyne*, Attorney General Eric Holder released a memo in August 2013 instructing US attorneys to decline to charge quantities necessary to trigger the mandatory minimum in cases with low-level and non-violent offenders who have little criminal history. The decrease in bunching could be a result of this memo and not the Supreme Court decision. To address that concern, I narrow the bandwidth of the RD design to 60 days before/after June 17th. Even then, I find a discontinuous decrease in bunching (although the standard errors are much larger).

[42] I do not conduct the traditional RD identifying assumption tests in this section. For one, the EOUSA data contain very few case-level covariates. Even more, the resulting discontinuity, whether it arises from prosecutors rushing to try cases before the Supreme Court decision or solely from prosecutors changing their behavior immediately after the decision, reveals that prosecutors were submitting evidence to judges that they believed would not hold up if submitted to a jury.

[43] I do not find a decrease in the fraction of cases recorded with 280-290g after the announcement that the Supreme Court would hear the case (in October 2012) or after the oral arguments (in January 2013). Unlike some Supreme Court cases, the ultimate ruling in June 2013 was not clear from the outset. At the time, the New York Times referred to the case as a "murky area of sentencing law" on which the Supreme Court had issued "contradictory rulings." For this reason, the announcement and the arguments alone would not provide sufficient evidence of whether the law would ultimately change.

2. Differential Costs of Acquiring Evidence by Race

In Section II, I introduce a model of prosecutor objectives that provides an explanation for bunching at mandatory minimum thresholds and differential bunching by race. In the model, differential bunching by race is generated by differential returns to sentencing a black versus a white offender. Furthermore, I assume that there is no difference in the cost of acquiring additional evidence by race. However, an alternative model that assumes no difference in the return to sentencing by race but a difference in the cost of acquiring evidence by race could also generate differential bunching by race.

Empirically, I cannot disentangle these two explanations if the cost of acquiring evidence by race is a characteristic that varies at the prosecutor level. However, if costs of acquiring evidence are constant within Census divisions or within Federal districts, then I can rule out that explanation. Using the nine Census region divisions, I find that the between-division standard deviation in the state-level racial disparity in bunching is equal to the within-division standard deviation in the state-level racial disparity in bunching. This is not possible if racial differences in the costs of acquiring evidence are constant within Census divisions. Furthermore, in Section V.C.3, I find that the within-district standard deviation in prosecutor-level bunching is similar to the between-district standard deviation in prosecutor-level bunching. Again, this is not possible if the costs of acquiring evidence are constant within Federal districts.

3. Heterogeneity by Offender Characteristics

Next, I estimate how bunching differs by various offender characteristics. Specifically, I estimate the following:

$$\begin{aligned}(Charged\,280-290g)_{it} = \alpha_0 + \beta_1(After2010 \times White)_{it} + \beta_2(After2010 \times Non-White)_{it} + \\ \beta_3(After2010 \times White \times Characteristic^H)_{it} + \beta_4(After2010 \times Non-White \times Characteristic^H)_{it} + \\ \beta_5 Characteristic^H_{it} + \beta_6 Non-White_{it} + \beta_5(Characteristic^H \times Non-White)_{it} + \epsilon_{it}\end{aligned} \tag{7}$$

where $Characteristic^H_{it}$ is a dummy variable representing the following offender characteristics: college education or more, male, above the median age for offenders, offense involves a weapon, above the median criminal history score, or above the median number of other current offenses. β_3 identifies the difference in bunching for defendant's with $Characteristic^H_{it} = 1$. I also estimate equation (6) with race interactions. This partially addresses concerns that white and black and Hispanic offender's are different on a wide range of other characteristics and that race may be a proxy for those characteristics. By estimating bunching by race and education, for example, I can compare black offenders with a college education to white offenders with a college education. If the racial disparity still exists within education categories, then this further suggests that the racial disparity is driven by attitudes about race. In Table 8, I show that the racial disparity in bunching exists even within all of these observably similar groups.

4. Heterogeneity by State-Level Racial Animus

One potential explanation of these results is that authorities believe that black and Hispanic drug offenders should be punished more harshly than white drug offenders. To explore this mechanism, I use a state-level measure of racial animus constructed by Stephens-Davidowitz (2014) based on intensity of Google searches including racial slurs in each state. Specifically, I match this measure to the USSC Sentencing data using the state of the federal district in which the offender is convicted. I take this measure of racial animus as a potentially valid measure of prosecutor tastes for several reasons: about half of government lawyers work in the same state they were born in (author's calculation from 2000 and 2010 publicly available Census samples), assistant US attorneys must reside in the district they serve in, and assistant US attorneys have a choice over where to apply. In addition, prosecutor decision-making happens in the "shadow of the law", in other words, the prosecutors decisions may be affected by the judge or potential jury a prosecutor would face in the district.[44]

I estimate the following to explore how bunching and the racial disparity in bunching differs by the amount of "racial animus" in the defendant's state of conviction:

$$(Charged\,280{-}290g)_{it} = \alpha_0 + \beta_1 After2010_{it} + \beta_2 RacialAnimus_{it} + \beta_3 (After2010 \times RacialAnimus)_{it} + \epsilon_{it} \tag{8}$$

where $RacialAnimus_{it}$ is a dummy variable equal to one if the state where the defendant is convicted is above the median on a measure of racial animus from Stephens-Davidowitz (2014) and equal to zero if it is below the median. β_3 identifies the difference in bunching after 2010 for states above the median. I also estimate equation (8) with race interactions as outlined in equation (7) above. If racial animus is correlated with some state-level preference for harsh sentencing, then I should find an effect for both white and black and Hispanic offenders. However, if the effect is driven by racist beliefs about black and Hispanic defendants, then it should only be present for those groups.

I find that in states with a higher level of racial animus, bunching at 280-290g is more prevalent and

[44] Recall that Alleyne v. US made the jury more important in mandatory minimum cases after 2013. This change led to stricted evidentiary standards for mandatory minimum cases (beyond a reasonable doubt versus preponderance of evidence). However, if juries are, on average, more racially biased than judges, then the effect of Alleyne v. US may be buffered by the increased racial bias of juries. I find that the fraction of cases at 280-290g in low racial animus states (below median) fell by 40% from 2011-2012 to 2014-2017. In high racial animus states (above median), the fraction of cases at 280-290g fell by 20%. This is suggestive evidence that Alleyne was, in fact, less effective in states with high racial animus. However, in all states, the increase in evidentiary standards led to a net decrease in cases at 280-290g.

this is only true for black and Hispanic offenders.[45] These results are in Table 9. Column 1 shows the main result from Table 1, and column 2 shows that the effect is larger in states above the median level of racial animus indicated by the Google index. Column 3 shows the main result by race from Table 2, and column 4 again shows that this effect is stronger in states designated as having more racial animus based on the Google index. Finally, this result is not driven by high racial animus states simply having a different racial composition of cases. I show in Table 8 that the racial disparity in bunching is similar in states and Federal districts that have a below median fraction of black and Hispanic offenders and in those that have an above median fraction of black and Hispanic offenders.

VI. Conclusion

For drug trafficking, a sharp jump in sentencing is triggered when an offense involves at or above a certain amount of drugs. In this paper, I show that there is substantial bunching at and above the point where the mandatory minimum sentence increases. Even more, that bunching is predominantly among black and Hispanic offenders and is concentrated in federal districts in states with high levels of racial animus.

Since the bunching only appears in prosecutor case management data and the final sentencing data but not in data on state-level convictions or drug seizures, it is likely a result of prosecutorial discretion. In fact, just 30% of attorneys account for 100% of the bunching observed in the case management data. In addition, bunching becomes less prevalent among prosecutors following a Supreme Court decision that requires stricter evidentiary standards for drug quantity evidence. Finally, the excess mass above the threshold appears to come from below the threshold, suggesting that prosecutors are shading the drug amounts upward to induce longer sentences.

Finally, the bunching in drug weights and the racial discrimination in bunching has meaningful sentencing consequences and implications for the racial sentencing gap. Depending on the counterfactual sentence imputed for the affected offenders, bunching at 280g can account for 2-7 percent of the racial disparity in crack-cocaine sentences. A highly conservative estimate suggests that being bunched at 280g adds 1-2 years to an offender's sentence. Multiple estimates suggest the cost of incarceration (combining direct care costs and the cost of lost current and future wages for the offender) is approximately $60,000 per person per year (Donohue and Sieglman 1998; Donohue 2009; Mueller-Smith 2015). Applying this cost to the 3.5% of crack-cocaine cases bunched at 280g from 2011-2014 implies a total cost of $16-$32 million. Assuming 3.5% of all drug cases from 1999-2014 were subject to similar discretion further implies a total cost of $1-$2 billion.

[45] Specifically, I split states by above/below the median racial animus. States above the median racial animus measure are: AL, AR, CT, DE, FL, GA, IL, IN, KY, LA, MD, MI, MO, MS, NC, NJ, NV, NY, OH, OK, PA, RI, SC, TN, and WV. States below the median racial animus measure are: AK, AZ, CA, CO, HI, IA, ID, KS, MA, ME, MN, MT, ND, NE, NH, NM, OR, SD, TX, UT, VA, VT, WA, WI, and WY.

References

Agan, Amanda and Sonja Starr. 2017. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *The Quarterly Journal of Economics* 133 (1): 191-235.

Anbarci, Nejat and Jungmin Lee. 2014. "Detecting Racial Bias in Speed Discounting: Evidence from Speeding Tickets in Boston." *International Review of Law and Economics* 38: 11–24.

Antonovics, Kate L. and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics* 91 (1): 163–177.

Anwar, Shamena and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (1): 127–151.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics* 127 (2): 1017–1055. https://doi.org/10.1093/qje/qjs014.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2017. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*.

Baker, Scott and Claudio Mezzetti. 2011. "Prosecutorial Resources, Plea Bargaining, and the Decision to Go to Trial." *The Journal of Law, Economics, and Organization* 17 (1): 149-167.

Bala, Nila. 2015. "Judicial Fact-Finding in the Wake of *Alleyne*." *New York University Review of Law & Social Change* 39 (1): 1–44.

Barkow, Rachel E. 2008. "Institutional Design and the Policing of Prosecutors: Lessons from Administrative Law." *Stanford Law Rev*iew 61: 869.

Becker, Gary S. 1957. *The Economics of Discrimination: An Economic View of Racial Discrimination*. University of Chicago.

Behaghel, Luc, and David M Blau. 2012. "Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age." *American Economic Journal: Economic Policy* 4 (4): 41–67. https://doi.org/10.1257/pol.4.4.41.

Berger v. United States, 295 U.S. 78. 1935.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem. 2015. "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy* 123 (6): 1311–55. https://doi.org/10.1086/683849.

Best, Michael and Henrik Jacobsen Kleven. 2018. "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK." Working paper.

Best, Michael, James Cloyne, Ethan Ilzetzki, and Henrik Jacobsen Kleven. 2018. "Interest Rates, Debt and Intertemporal Allocation: Evidence from Notched Mortgage Contracts in the United Kingdom." Working paper.

Bibas, Stephanos. 2004. "Plea Bargaining Outside the Shadow of Trial." *Harvard Law Review* 2463–2547.

Bibel, Daniel. 2015. "Considerations and Cautions Regarding NIBRS Data: A View From the Field." *Justice Research and Policy* 16 (2): 185-194.

Bjerk, David. 2005. "Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion under Mandatory Minimum Sentencing." *The Journal of Law and Economics* 48 (2): 591–625.

Bjerk, David. 2007. "Guilt Shall Not Escape or Innocence Suffer? The Limits of Plea Bargaining When Defendant Guilt Is Uncertain." *American Law and Economics Review* 9 (2): 305–329. https://doi.org/10.1093/aler/ahm010.

Bjerk, David. 2017. "Mandatory Minimums and the Sentencing of Federal Drug Crimes." *Journal of Legal Studies* 46 (1): 93-128

Boylan, Richard T. 2005. "What Do Prosecutors Maximize? Evidence from Careers of U.S. Attorneys." *American Law and Economics Review* 7 (2): 379-402.

Brown, Kristine M. 2013. "The Link between Pensions and Retirement Timing: Lessons from California Teachers." *Journal of Public Economics* 98 (February): 1–14.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2018. "The Effect of Minimum Wages on Low-Wage Jobs: Evidence from the United States Using a Bunching Estimator." Working paper.

Chen, Zhao, Zhikuo Liu, Juan Carlos Suarez Serrato, and Daniel Yi Xu. 2018. "Notching R&D Investment with Corporate Income Tax Cuts in China." Working paper.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103 (7): 2683–2721.

Dee, Thomas S., Will Dobbie, Brian A. Jacob, and Jonah Rockoff. 2017. "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations." Working paper.

Diamond, Rebecca and Petra Persson. 2017. "The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests." Working paper.

Donohue, John J., III, and Peter Siegelman. 1998. "Allocating Resources among Prisons and Social Programs in the Battle against Crime." *Journal of Legal Studies* 27 (1): 1–43.

Donohue, John J., III. 2009. "Assessing the Relative Benefits of Incarceration: Overall Changes and the Benefts on the Margin." *Do Prisons make Us Safer? The Benefts and costs of the Prison Boom*, edited by Steven Raphael and Michael A. Stoll, 269–342. New York: Russell Sage Foundation

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy* 8 (3): 203–32. https://doi.org/10.1257/pol.20150083.

Easterbrook, Frank H. 1983. "Criminal Procedure as a Market System." *The Journal of Legal Studies* 12 (2): 289–332.

Fack, Gabrielle, and Camille Landais. 2016. "The Effect of Tax Enforcement on Tax Elasticities: Evidence from Charitable Contributions in France." *Journal of Public Economics* 133 (January): 23–40. https://doi.org/10.1016/j.jpubeco.2015.10.004.

Friedberg, Leora. 2000. "The Labor Supply Effects of the Social Security Earnings Test." *Review of Economics and Statistics* 82 (1): 48–63. doi:http://www.mitpressjournals.org/loi/rest.

Gelber, Alexander M., Damon Jones, and Daniel W. Sacks. 2013. "Estimating Earnings Adjustment Frictions: Method and Evidence from the Social Security Earnings Test." NBER working paper.

Glaeser, Edward L., Daniel P. Kessler, and Anne Morrison Piehl. 2000. "What Do Prosecutors Maximize? An Analysis of Federalization of Drug Crimes." *American Law and Economics Review* 2 (2): 259-290.

Goncalves, Felipe and Steven Mello. 2018. "A Few Bad Apples? Racial Bias in Policing." Working paper.

Gordon, Sanford C. 2009. "Assessing Partisan Bias in Federal Public Corruption Prosecutions." *American Political Science Review* 103 (04): 534–54. https://doi.org/10.1017/S0003055409990207.

Grogger, Jeffrey and Greg Ridgeway. 2006. "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness." *Journal of the American Statistical Association* 101 (475): 878– 887.

Grossman, Gene M. and Michael L. Katz. 1983. “Plea Bargaining and Social Welfare.” *American Economic Review* 73 (4): 749-757.

Hansen, Benjamin. 2015. “Punishment and Deterrence: Evidence from Drunk Driving.” *American Economic Review* 105 (4): 1581–1617. doi:10.1257/aer.20130189.

Honold, Dan. 2014. “Note: Quantity, Role, and Culpability in the federal Sentencing Guidelines.” *Harvard Journal on Legislation* 51: 389-314.

Imbens, Guido, and Karthik Kalyanaraman. 2012. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *Review of Economic Studies* 79(3): 933–959.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Soren Pedersen, and Emmanuel Saez. “Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark.” 2011. *Econometrica* 79 (3): 651–692. https://doi.org/10.3982/ECTA9113.

Kleven, Henrik Jacobsen and Mazhar Waseem. 2013. “Using notches to uncover optimization frictions and structural elasticities: theory and evidence from Pakistan.” *The Quarterly Journal of Economics* 128: 669–723

Kleven, Henrik Jacobsen. 2016. “Bunching.” *Annual Review of Economics* 8: 435–464.

Knorre, Alexey. 2017. “How to detect fraud in drug crime records? An analysis of 585,000 crime records in Russia.” Presented at 17th Annual Conference of the ESC: EuroCrim 2017.

Knowles, John, Nicola Persico, and Petra Todd. 2001. “Racial Bias in Motor Vehicle Searches: Theory and Evidence.” *Journal of Political Economy* 109 (1): 203–229.

Landes, William M. 1971. “An Economic Analysis of the Courts.” *Journal of Law and Economics* 14 (1): 61-107.

Lynch, Mona. *Hard Bargains: The Coercive Power of Drug Laws in federal Court.* New York: Russell Sage Foundation, 2016.

Mueller-Smith, Michael. 2015. “The Criminal and Labor Market Impacts of Incarceration.” University of Michigan Working Paper.

Mustard, David B. 2001. “Racial, Ethnic and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts.” *Journal of Law and Economics* 44: 285–314.

Nyhan, Brendan and M. Marit Rehavi. 2017. “Tipping the Scales? Testing for Political Influence on Public Corruption Prosecutions.” Working paper.

Rehavi, M. Marit, and Sonja B. Starr. 2014. “Racial Disparity in federal Criminal Sentences.” *Journal of Political Economy* 122 (6):1320–54. https://doi.org/10.1086/677255.

Reinganum, Jennifer F. 1988. “Plea Bargaining and Prosecutorial Discretion.” *American Economic Review* 78 (4): 713-728.

Rhodes, William, Ryan Kling, Jeremy Luallen, and Christina Dyous. 2015. “Federal Sentencing Disparity: 2005-2012.” Bureau of Justice Statistics Working Paper Series.

Saez, Emmanuel. 2010. “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy* 2 (3): 180–212. doi:10.1257/pol.2.3.180.

Sallee, James M., and Joel Slemrod. 2012. “Car Notches: Strategic Automaker Responses to Fuel Economy Policy.” *Journal of Public Economics* 96 (11): 981-999.

Shermer, Lauren O’Neill and Brian D. Johnson. 2009. “Criminal Prosecutions: Examining Prosecutorial Discretion and Charge Reductions in U.S. Federal District Courts.” *Justice Quarterly* 1-37.

Shively, Michael. 2005. "Study of Literature and Legislation on Hate Crime in America." National Insitute of Justice.

Simon, Jonathan. 2007. *Governing Through Crime: How the War on Crime and American Democracy Created a Culture of Fear*. Oxford: Oxford Univ. Press

Sklansky, David Alan. 2018. "The Problems with Prosecutors." *Annual Review of Criminology* 1: 451–469.

Spohn, Cassia, John Gruhl, and Susan Welch. 1987. "The Impact of the Ethnicity and Gender of Defendants on the Decision to Reject or Dismiss Felony Charges." *Criminology* 25 (1): 175–192.

Stephens-Davidowitz, Seth. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence using Google Search Data." *Journal of Public Economics* 118: 26-40.

Sterling, Eric. 1999. "Drug Laws and Snitching: A Primer." PBS: Frontline. http://www.pbs.org/wgbh/pages/frontline/shows/snitch/

Stuntz, William J. 2004. "Plea Bargaining and Criminal Law's Disappearing Shadow." *Harvard Law Review* 117 (8): 2548. https://doi.org/10.2307/4093405.

Traxler, Christian, Franz G. Westermaier, and Ansgar Wohlschlegel. 2018. "Bunching on the Autobahn? Speeding Responses to a 'Notched' Penalty Scheme." *Journal of Public Economics* 157: 78–94.

United States Department of Justice (DOJ). 2002. "Federal Cocaine Offenses: An Analysis of Crack and Powder Penalties." https://www.justice.gov/archive/olp/pdf/crack_powder2002.pdf

United States Department of Justice (DOJ). 2011. "The Department of Justice's International Prisoner Transfer Program." https://oig .justice.gov/reports/2011/e1202.pdf.

United States Sentencing Commission (USSC), 2015a. "Report to the Congress: Impact of the Fair Sentencing Act of 2010." https://www.ussc.gov/sites/default/files/pdf/news/congressional-testimony-and-reports/drug-topics/201507_RtC_Fair-Sentencing-Act.pdf

United States Sentencing Commission (USSC). 2015b. "United States Sentencing Comission (USSC) Guidelines Manual." https://www.ussc.gov/sites/default/files/pdf/guidelines-manual/2015 /GLM-Full.pdf

United States Sentencing Commission (USSC). 2017. "Demographic Differences in Sentencing: An Update to the 2012 Booker Report." https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2017/20171114_Demographics.pdf

West, Jeremy. 2018. "Racial Bias in Police Investigations." Working paper.

Wright, Ronald. 2006. "Federal or State? Sorting as a Sentencing Choice." *Criminal Justice* 21 (2): 16-21.

Yang, Crystal S. 2016. "Resource Constraints and the Criminal Justice System: Evidence from Judicial Vacancies." *American Economic Journal: Economic Policy* 8 (4): 289–332.

Zaresani, Arezou. 2016. "Adjustment Costs and Incentives to Work: Evidence from a Disability Insurance Program." University of Calgary working paper.

Tables and Figures

Table 1. Summary Statistics for USSC Sentencing Data.

	1999-2010	2011-2015
Black or Hispanic	0.921	0.939
	(0.270)	(0.239)
Age	31.184	34.169
	(8.512)	(8.748)
Male	0.915	0.917
	(0.279)	(0.277)
College or more	0.127	0.148
	(0.332)	(0.356)
High school or more	0.509	0.598
	(0.500)	(0.490)
Not US citizen	0.046	0.033
	(0.209)	(0.178)
Weapon involved	0.262	0.297
	(0.440)	(0.457)
Number of counts	1.606	1.725
	(1.428)	(1.741)
Only one drug charged	0.694	0.487
	(0.461)	(0.500)
Drug wseight (in grams)	105.791	121.062
	(165.717)	(179.802)
Sentence (in years)	9.315	7.819
	(7.102)	(5.840)
Observations	47,612	9,489

Notes. The table above describes defendants found in the USSC sentencing data pre- and post-2010. The mean value of each variable is reported with standard deviations in parentheses. The statistics above reflect data cleaning in which the following cases are removed: cases with missing drug weight values (including those cases coded as a range), cases with reported problems in the drug weight variables, cases where judges change or do not accept the findings of fact for drug weights, cases at and above 1000g.

Table 2. Effect of Changing Threshold on Bunching at 280-290g.

	Pr(280-290g Crack-Cocaine Recorded)			
	(1)	(2)	(3)	(4)
After 2010	0.0347***			0.0765***
	(0.00204)			(0.0138)
After 2010 x White		0.0125**	0.0130**	
		(0.0053)	(0.0053)	
After 2010 x Non-White		0.0360***	0.0326***	
		(0.0021)	(0.0022)	
Constant	0.0051***	0.0032***	0.0027**	0.0068**
	(0.0003)	(0.0010)	(0.0011)	(0.0031)
P-value: W (White) = NW (Non-White)	-	0.0000	0.0004	-
Hispanic Offenders Excluded	No	No	Yes	No
Jury Trials Only	No	No	No	Yes
Observations	57,101	52,940	47,932	2,706
R-squared	0.015	0.016	0.014	0.088

Notes. Robust standard errors in parentheses. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams. The row "P-value: W (White) = NW (Non-White)" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." In the remaining tables, I abbreviate the label to "P-value: W= NW." Specifications with the white/non-white and after 2010 interactions also include a dummy variable equal to one for black and Hispanic offenders. Column 4 reports the main bunching result for cases that go to trial by jury–there are zero cases for white offenders with 280-290g in this category. "Non-white" refers to black and Hispanic offenders–in many cases, I will refer to "black and Hispanic offenders" directly, but for brevity, I refer to these offenders as "non-white" in the tables.
*** p<0.01, ** p<0.05, * p<0.1

Table 3a. Missing Mass in the Distribution of Drug Amounts, with Linear Trend Interaction and State FEs

Panel A. Analysis of Changes in the 0-100g Range.

	Pr(0-5g)	Pr(5-28g)	Pr(28-50g)	Pr(50-60g)	Pr(60-100g)
	(1)	(2)	(3)	(4)	(5)
After 2010 x White	0.0378	-0.0666**	0.0056	-0.0215	-0.0142
	(0.0272)	(0.0305)	(0.0226)	(0.0183)	(0.0218)
After 2010 x Non-White	0.0024	-0.0907***	0.0296***	-0.0077	-0.0034
	(0.0059)	(0.0081)	(0.0065)	(0.0050)	(0.0063)
Constant	0.1384***	0.2820***	0.1106***	0.0880***	0.1246***
	(0.0133)	(0.0166)	(0.0109)	(0.0095)	(0.0111)
P-value: W = NW	0.2030	0.4458	0.3074	0.4694	0.6337
Observations	52,530	52,530	52,530	52,530	52,530
R-squared	0.032	0.019	0.006	0.007	0.007

Panel B. Analysis of Changes in the 100-1000g Range.

	Pr(100-280g)	Pr(280-290g)	Pr(290-470g)	Pr(470-600g)	Pr(600-1000g)
	(6)	(7)	(8)	(9)	(10)
After 2010 x White	0.0185	0.0164**	0.0019	0.0182*	0.0040
	(0.0270)	(0.0083)	(0.0144)	(0.0110)	(0.0137)
After 2010 x Non-White	0.0094	0.0345***	0.0117***	0.0062**	0.0080**
	(0.0076)	(0.0033)	(0.0041)	(0.0029)	(0.0033)
Constant	0.1642***	0.0043**	0.0454***	0.0128***	0.0297***
	(0.0132)	(0.0020)	(0.0073)	(0.0044)	(0.0056)
P-value: W = NW	0.7458	0.0415	0.5101	0.2916	0.7778
Observations	52,530	52,530	52,530	52,530	52,530
R-squared	0.012	0.023	0.007	0.006	0.010

Notes. Robust standard errors in parentheses. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams, and all specifications include state fixed effects and a linear time trend. In addition, specifications in this table include an interaction between the After 2010 dummy and the linear time trend. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." Panel A displays analysis of changes from 0-100g and Panel B displays analysis of changes from 100-1000g.
*** p<0.01, ** p<0.05, * p<0.1

Table 3b. Missing Mass in the Distribution of Drug Amounts, with Various Time Trend Controls and State FEs

	Pr(< 280g)	Pr(280-290g)	Pr(> 290g)
	(1)	(2)	(3)
Panel A. No Interaction with Time Trend			
After 2010 x White	-0.0672***	0.0120**	0.0552***
	(0.0151)	(0.0055)	(0.0143)
After 2010 x Non-White	-0.0605***	0.0344***	0.0262***
	(0.0051)	(0.0023)	(0.0047)
Constant	0.9372***	0.0059***	0.0569***
	(0.0053)	(0.0013)	(0.0052)
Observations	52,530	52,530	52,530
R-squared	0.023	0.023	0.020
P-value: W = NW	0.6624	0.0001	0.0435
Panel B. Interaction with Linear Time Trend			
After 2010 x White	-0.0404*	0.0164**	0.0240
	(0.0229)	(0.0083)	(0.0218)
After 2010 x Non-White	-0.0604***	0.0345***	0.0259***
	(0.0064)	(0.0033)	(0.0057)
Constant	0.9078***	0.0043**	0.0879***
	(0.0100)	(0.0020)	(0.0099)
Observations	0.3996	0.0415	0.9330
R-squared	52,530	52,530	52,530
P-value: W = NW	0.024	0.023	0.020
Panel C. Interaction with Quadratic Time Trends			
After 2010 x White	0.0028	0.0133	-0.0161
	(0.0303)	(0.0099)	(0.0291)
After 2010 x Non-White	-0.0259***	0.0302***	-0.0042
	(0.0085)	(0.0040)	(0.0078)
Constant	0.8786***	0.0038	0.1176***
	(0.0192)	(0.0040)	(0.0188)
Observations	52,530	52,530	52,530
R-squared	0.025	0.023	0.021
P-value: W = NW	0.3612	0.1141	0.6940

Notes. Robust standard errors in parentheses. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams, and all specifications include state fixed effects. Panel A results include a linear time trend, Panel B results include a linear time trend and the interaction between the trend and the After 2010 dummy, and Panel C results include a quadratic time trend and the interaction between the trend and the After 2010 dummy. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White."
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Sentencing Consequences of Being Above the Threshold Amount

	Years Sentenced					
	(1)	(2)	(3)	(4)	(5)	(6)
Above 280g	-0.580**	0.0621			0.00410	-0.0576
	(0.289)	(0.691)			(0.294)	(0.461)
Above 280g x After 2010	2.332***	2.181**			0.971*	2.836***
	(0.508)	(1.102)			(0.535)	(0.842)
Above 50g			0.755***	0.955***	1.469***	2.101***
			(0.128)	(0.158)	(0.180)	(0.227)
Above 50g x After 2010			-1.387***	-1.063***	-1.298***	-2.058***
			(0.270)	(0.357)	(0.451)	(0.445)
Constant	12.93***	11.48***	9.664***	9.540***	13.12***	14.08***
	(0.170)	(0.565)	(0.114)	(0.116)	(3.298)	(3.709)
Bandwidth	±250g	±50g	±250g	±50g	±250g	±250g
Includes Life & <1 Month	No	No	No	No	No	Yes
Observations	29,767	2,800	49,154	14,713	29,064	31,134
R-squared	0.037	0.015	0.070	0.035	0.038	0.031

Notes. Robust standard errors in parentheses. The specifications above estimate regression discontinuity style models, and thus, they all include a control for the running variable (amount of drugs centered at the appropriate mandatory minimum threshold) and an interaction between Above 280g (or Above 50g) and the running variable. In addition, all specifications above include a time trend to capture the gradual decline in sentences over time. Column 6 includes life sentences (coded as 70 years) and sentences less than 1 month (coded as 0 years).

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

Table 5. Bunching Analysis for Potential Mechanisms

Panel A. Analysis of Bunching in State Courts and in Drug Seizures

	Pr(200-400g)	Pr(200-400g)	Pr(280-290g)	Pr(280-290g)	Pr(280-290g)
	(1)	(2)	(3)	(4)	(5)
After 2010	0.00005		-0.0002***		-0.0006***
	(0.0005)		(.0001)		(0.0002)
After 2010 x White		0.0004		-0.0001	
		(0.0011)		(0.0001)	
After 2010 x Non-White		0.0002		-0.0003***	
		(0.0005)		(0.0001)	
Constant	0.0051***	0.0085***	0.0004***	0.0002***	0.0010***
	(0.0003)	(0.0005)	(0.00005)	(0.0001)	(0.0001)
Data Analyzed	FL Convictions	FL Convictions	Drug Seizures, NIBRS	Drug Seizures, NIBRS	Drug Evidence, DEA STRIDE
Drugs Included	Cocaine, all types	Cocaine, all types	Crack-cocaine	Crack-cocaine	Cocaine, all types
P-value: W = NW	-	0.8148	-	0.2537	-
Observations	214,573	214,573	203,532	188,737	100,306
R-squared	0.000	0.001	0.000	0.000	0.0000

Panel B. Analysis of Bunching in Prosecutor Case Files and Final Sentencing

	Pr(280-290g)	Pr(200-400g)	Pr(200-400g)	Pr(280-290g)	Pr(280-290g)
	(6)	(7)	(8)	(9)	(10)
After 2010	0.0777***	0.0362***		0.0347***	
	(0.0056)	(0.0119)		(0.00204)	
After 2010 x White			0.00521		0.0125**
			(0.0278)		(0.0053)
After 2010 x Non-White			0.0394***		0.0360***
			(0.0122)		(0.0021)
Constant	0.0039***	0.1060***	0.117***	0.0051***	0.0032***
	(0.0005)	(0.00695)	(0.0148)	(0.0003)	(0.0010)
Data Analyzed	EOUSA Case Management System	USSC Sentencing, FL only	USSC Sentencing, FL only	USSC Sentencing	USSC Sentencing
Drugs Included	Crack-cocaine	Cocaine, all types	Cocaine, all types	Crack-cocaine	Crack-cocaine
P-value: W = NW	-	-	0.2211	-	0.0000
Observations	19,363	7,178	7,178	57,101	52,940
R-squared	0.049	0.001	0.002	0.015	0.016

Notes. Robust standard errors in parentheses. When possible, the specifications above use a sample of offenses with drug amounts between 0 grams and 1000 grams. Analyses of state-level drug convictions do not make this restriction since the state reports broad drug weight categories instead of specific amounts. When broad categories are analyzed, a linear trend in year is included. The row "P-value: W= NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." In Panel A: columns 1-2 show an analysis of reported drug amounts for state-level drug convictions in Florida, columns 3-4 show an analysis of weights for seized drugs reported to the FBI through the National Incident Based Reporting System, and column 5 shows an analysis of weights for drugs sent to DEA laboratories. In Panel B: column 6 shows an analysis of weights recorded in case management files from the Executive Office of the US attorney, columns 7-8 show an analysis of weights from USSC sentencing data for federal convictions in FL using broad drug categories and all types of cocaine, and columns 9-10 show the main bunching results from Table 2 for all federal crack-cocaine convictions in the USSC sentencing data.
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Change in Bunching by Prosecutors after *Alleyne v. United States* Decision

	Pr(Case Recorded with 280-290g)			
	(1)	(2)	(3)	(4)
After June 17th, 2011-2016	0.0070	-0.0049	0.0041	-0.0206
	(0.0260)	(0.0284)	(0.0295)	(0.0406)
After June 17th, 2013	-0.1740**	-0.1518*	-0.1433	-0.1289
	(0.0813)	(0.0920)	(0.0935)	(0.1246)
Constant	0.1620	0.1626	0.1576	0.2093
	(0.1520)	(0.1519)	(0.1520)	(0.1776)
Bandwidth	±150 days	±130 days	±120 days	±60 days
Observations	1,937	1,672	1,513	754
R-squared	0.009	0.010	0.009	0.013

Notes. Standard errors clustered at the date the case is received in parentheses. The specifications above estimate regression discontinuity style models, and thus, they all include a control for the running variable (number of days from June 17th in a given year) and an interaction between After June 17th and the running variable. In addition, all specifications above include day-of-week fixed effects. The ±130 day bandwidth is selected from the Imbens-Kalyanaraman optimal bandwidth procedure for the year 2013.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7a. Offender Drug-Holding Behavior by Race.

	Weight	Weight	Pr(280-290g)
	(1)	(2)	(3)
Black	1.466***	-0.300	0.0001
	(0.269)	(0.592)	(0.0001)
Constant	10.65***	19.47***	0.0003**
	(0.398)	(0.818)	(0.0001)
Excluding Possession <= 1g	No	Yes	No
Observations	188,737	98,258	188,737
R-squared	0.004	0.008	0.000
P-value: W = NW	-	-	-

Notes. Robust standard errors in parentheses. This analysis uses the weights of seized drugs reported to the FBI through the National Incident Based Reporting System. Columns 1 and 2 show the relationship between race of offender and drug weight seized. Column 3 shows the relationship between race of offender and probability the amount seized is between 280-290g. All specifications include state fixed effects and controls for age and sex.

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

Table 7b. Offender Drug-Holding Behavior by Race, After Fair Sentencing Act in 2010

	Weight	Pr(0-5g)	Pr(5-28g)	Pr(28-50g)	Pr(50-280g)	Pr(270-280g)	Pr(280-290g)	Pr(>290g)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After 2010 x White	-0.484	0.0327***	-0.0251***	-0.00225	-0.00615***	-1.15e-05	-5.02e-05	0.000900
	(0.592)	(0.00433)	(0.00389)	(0.00172)	(0.00136)	(2.27e-05)	(0.000127)	(0.000752)
After 2010 x Black	-2.824***	0.0539***	-0.0274***	-0.00829***	-0.0163***	-0.000172***	-0.000238**	-0.00160***
	(0.301)	(0.00308)	(0.00284)	(0.00118)	(0.00103)	(3.80e-05)	(0.000103)	(0.000395)
Constant	10.50***	0.725***	0.200***	0.0364***	0.0345***	0.000102	0.000243**	0.00431***
	(0.417)	(0.00381)	(0.00341)	(0.00155)	(0.00147)	(6.23e-05)	(0.000114)	(0.000583)
Observations	188,737	188,737	188,737	188,737	188,737	188,737	188,737	188,737
R-squared	0.004	0.028	0.021	0.004	0.008	0.000	0.000	0.001
P-value: W = NW	0.00046	0.00006	0.64257	0.00377	0.00000	0.00045	0.25161	0.00357

Notes. Robust standard errors in parentheses. This analysis uses the weights of seized drugs reported to the FBI through the National Incident Based Reporting System. Column 1 shows how the weight of an offender's seized drugs changes by race after 2010. Columns 2-8 show how the probability an offender's seized drugs are in a certain bin changes by race after 2010. All specifications include state fixed effects and controls for age and sex.
*** p<0.01, ** p<0.05, * p<0.1

Table 8. Degree of Bunching Post-2010 by Race and Offender Characteristics.

	Pr(280-290g)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After 2010 x White	0.0171**	0.0065	0.0143	0.0129**	0.0160**	0.0103*	0.0149**
	(0.0068)	(0.0063)	(0.0087)	(0.0062)	(0.0071)	(0.0058)	(0.0068)
After 2010 x Non-White	0.0363***	0.0235***	0.0424***	0.0303***	0.0452***	0.0306***	0.0471***
	(0.0023)	(0.0072)	(0.0037)	(0.0024)	(0.0036)	(0.0025)	(0.0037)
After 2010 x White x Char.	-0.0207***	0.0109	-0.0024	-0.0015	-0.0095	0.0089	-0.0074
	(0.0072)	(0.0100)	(0.0109)	(0.0120)	(0.0107)	(0.0135)	(0.0104)
After 2010 x Non-White x Char.	-0.0042	0.0131*	-0.0102**	0.0191***	-0.0163***	0.0157***	-0.0189***
	(0.0061)	(0.0076)	(0.0046)	(0.0051)	(0.0044)	(0.0047)	(0.0045)
Constant	0.0032***	0.0022	0.0031**	0.0033***	0.0013*	0.0036***	0.0031***
	(0.0011)	(0.0015)	(0.0014)	(0.0011)	(0.0008)	(0.0012)	(0.0012)
Characteristic	College	Male	Above Med. Age	Weapon	Above Med. Crim. Hist. Points	Above Med. # of Other Counts	State Above Med. % of Non-White Cases
P-value: W = NW	0.0074	0.0764	0.0031	0.0085	0.0002	0.0012	0.0000
P-value: W+Char. = NW+Char.	0.0000	0.0177	0.0043	0.0007	0.0078	0.0352	0.0121
Observations	52,389	49,049	52,712	52,233	52,725	52,742	52,745
R-squared	0.016	0.016	0.017	0.017	0.017	0.017	0.017

Notes. Robust standard errors in parentheses. “Characteristic” or “Char.” represents a dummy variable that is an offender or case characteristic. The specific offender characteristic of interest is noted in the “Characteristic” row. For example, when the “Characteristic” is “College”, then “Characteristic” is equal to one if the offender’s educational attainment is college or more and is equal to zero if the offender’s educational attainment is less than college. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams. The row “P-value: W = NW” reports the p-value from a test of the null hypothesis that the coefficient on “After 2010 x White” is equal to the coefficient on “After 2010 x Non-White.” The row “P-value: W+Char. = NW+Char.” reports the p-value from a test of the null hypothesis that the combined coefficients on “(After 2010 x White)+(After 2010 x White x Characteristic)” is equal to the combined coefficients on “(After 2010 x Non-White)+(After 2010 x Non-White x Characteristic).” Male is equal to one if the offender is male and equal to zero if not. Above median age is equal to one if the offender is above the median age for offenders and equal to zero if not. Weapon is equal to one if the offense involves a weapon and equal to zero if not. Above median crim. hist. points is equal to one if the offender has a criminal history score above the median criminal history score for offenders and equal to zero if not. Above the median # of other counts is equal to one if the offender has above the median number of other criminal counts for offenders and equal to zero if not. The final column examines differences in bunching for states with above/below the median fraction of non-white cases.

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

Table 9. Differential Bunching by State-Level Racial Animus

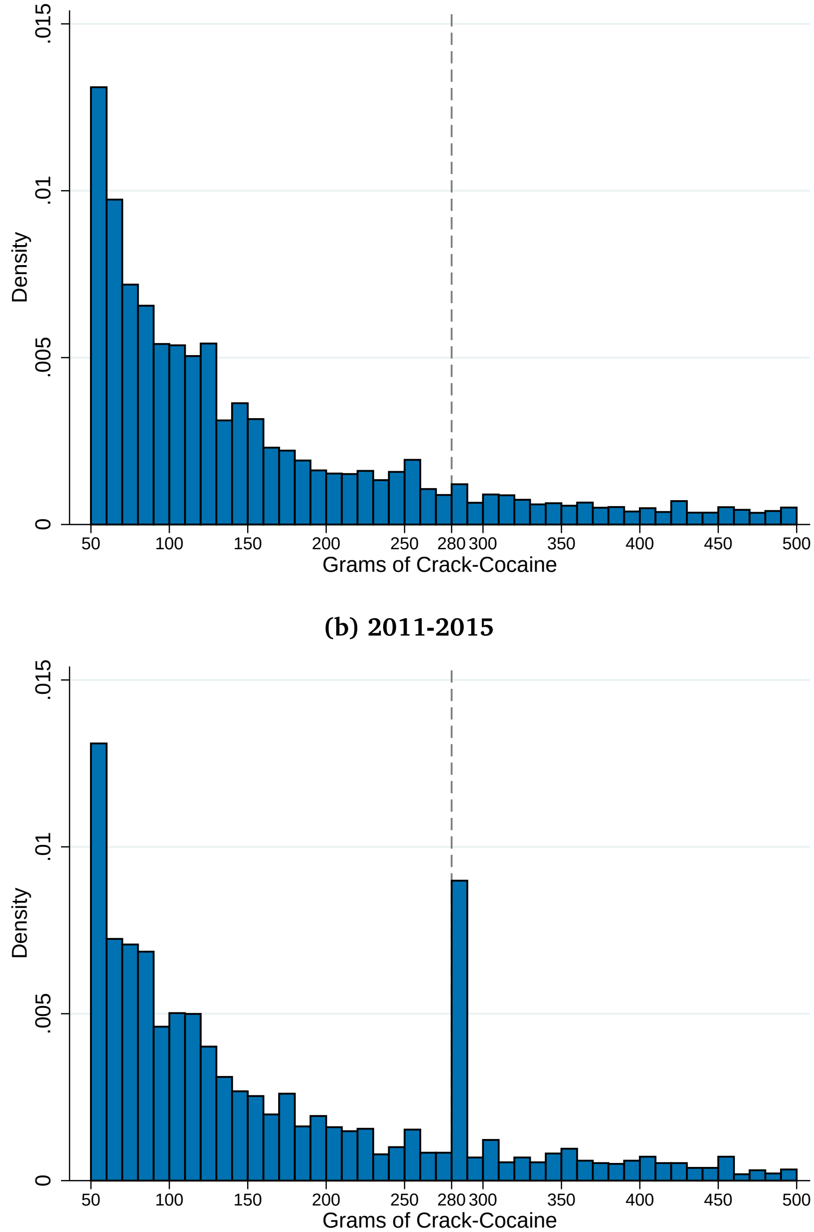
	Pr(280-290g Crack-Cocaine)			
	(1)	(2)	(3)	(4)
After 2010	0.0347***	0.0151***		
	(0.0020)	(0.00294)		
After 2010 x White			0.0125**	0.0085
			(0.0053)	(0.0081)
After 2010 x Non-White			0.0360***	0.0156***
			(0.0021)	(0.0031)
State > Median Racial Animus		-0.0006		-0.0030
		(0.0007)		(0.0023)
After 2010 x Racial Animus		0.0243***		
		(0.00388)		
After 2010 x White x				0.0067
Racial Animus				(0.0108)
After 2010 x Non-White x				0.0250***
Racial Animus				(0.0041)
Constant	0.0051***	0.0056***	0.0032***	0.0052**
	(0.0003)	(0.0006)	(0.0010)	(0.0021)
P-value: W = NW	-	-	0.0000	0.4142
P-value: W+RA = NW+RA	-	-	-	0.0008
Observations	57,101	55,734	52,940	51,679
R-squared	0.015	0.016	0.016	0.017

Notes. Robust standard errors in parentheses. "State Above Median Racial Animus" and "Racial Animus" represent a dummy variable that is equal to one if the state where the offender is convicted is above the median on a measure of racial animus from Stephens-Davidowitz (2014) and is equal to zero if the state is below the median. The measure is based on the intensity of Google searches in each state that involve racial slurs. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." The row "P-value: W+RA = NW+RA" reports the p-value from a test of the null hypothesis that the combined coefficients on "(After 2010 x White)+(After 2010 x White x Racial Animus)" is equal to the combined coefficients on "(After 2010 x Non-White)+(After 2010 x Non-White x Racial Animus)."

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

Figure 1. Changing Distribution of Drug Amounts Around 280g Pre- and Post-2010.

(a) 1999-2010

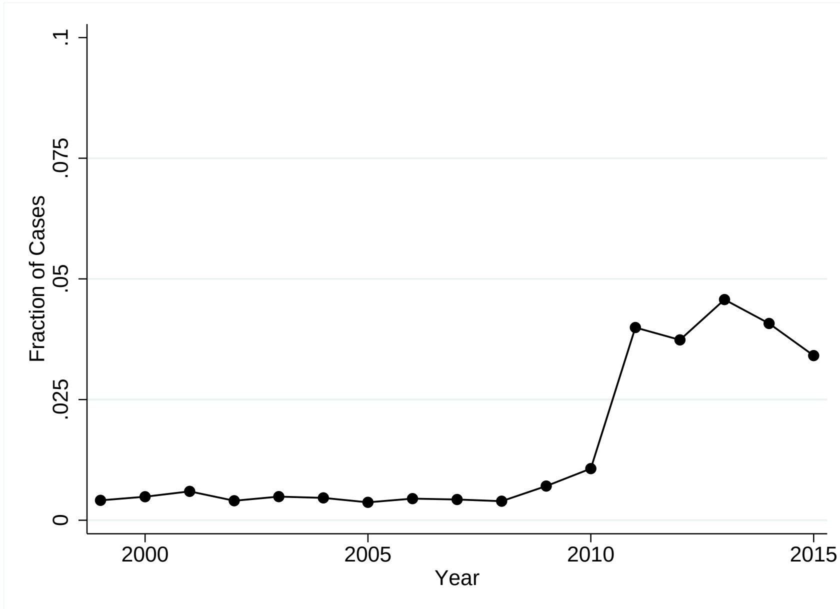


(b) 2011-2015

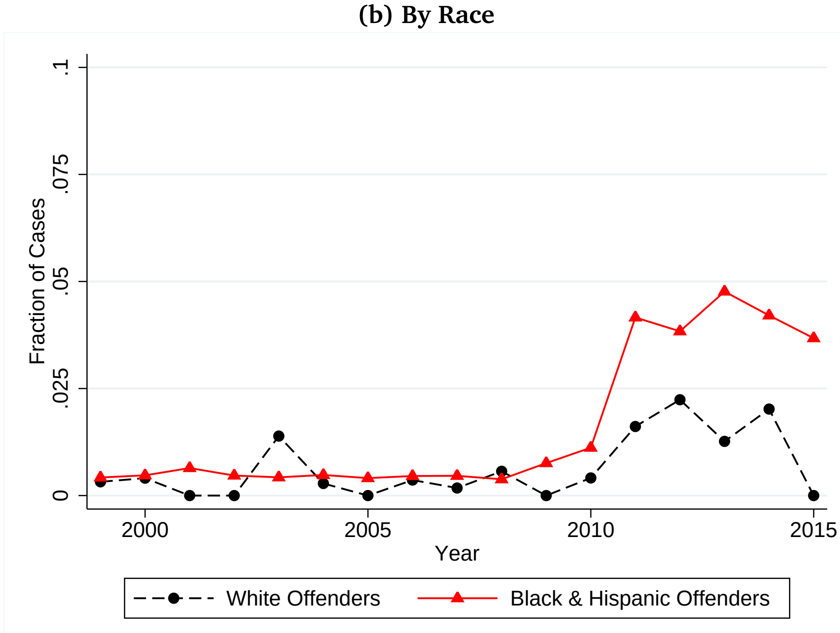
Notes. Both figures plot the distribution of drug amounts recorded in federal crack-cocaine sentences starting at 50 grams and ending at 500 grams. Panel (a) displays this distribution for cases sentenced from 1999-2010, when the mandatory minimum threshold was 50 grams. Panel (b) displays this distribution for cases sentenced from 2011-2015, when the mandatory minimum threshold was 280 grams.

Figure 2. Changing Fraction of Cases with 280-290g Over Time.

(a) All Offenders



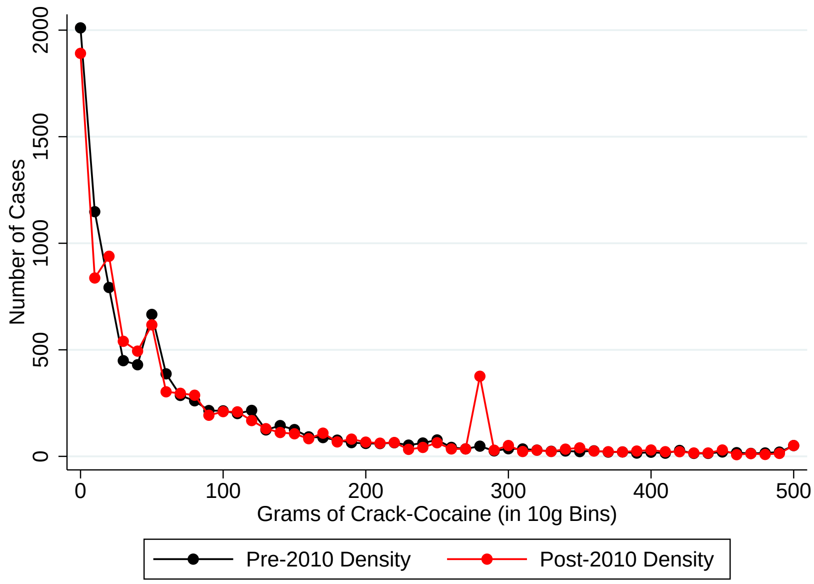
(b) By Race



Notes. Both figures plot the fraction of cases recorded with 280-290 grams of crack-cocaine by year. The denominator is all crack-cocaine cases under 1000 grams. Panel (a) displays this fraction over time for all offenders. Panel (b) displays this fraction over time for white and black and Hispanic offenders.

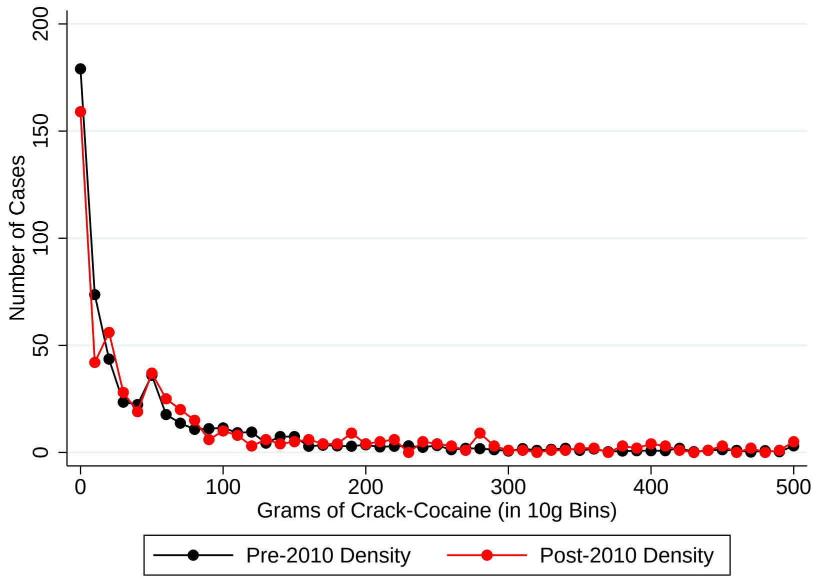
Figure 3. Pre-2010 Density and Post-2010 Density

(a) All Offenders

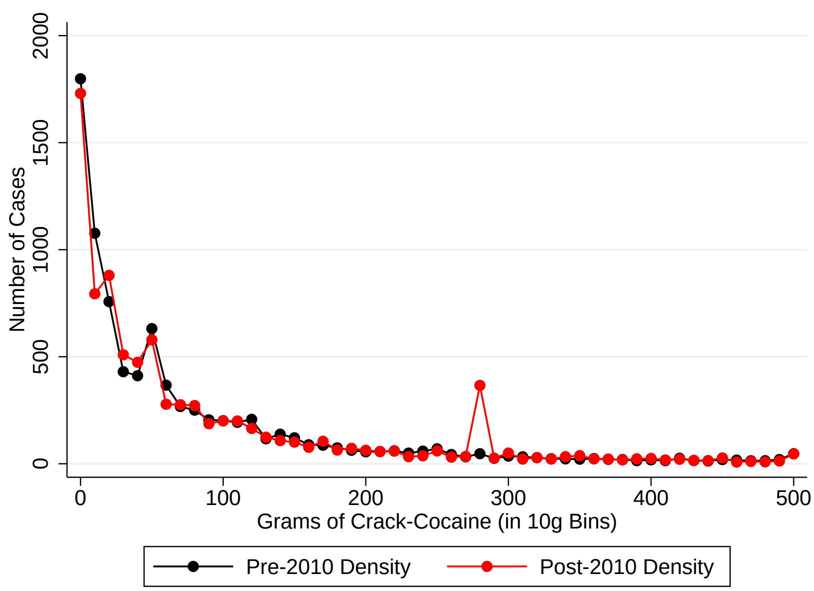


Notes. The figure above plots the scaled density of drug quantities pre-2010 (in black) and the actual density of drug quantities post-2010 (in red). The amounts are aggregated into 10-gram bins and limited to drug quantities under 1000g.

(b) White Offenders

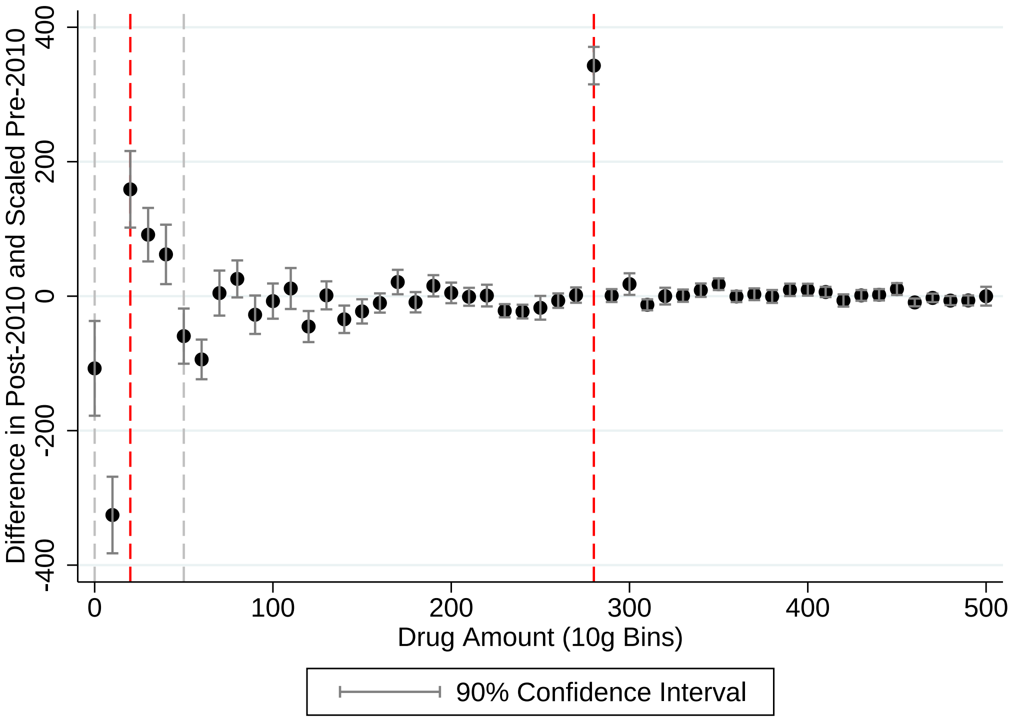


(c) Black and Hispanic Offenders



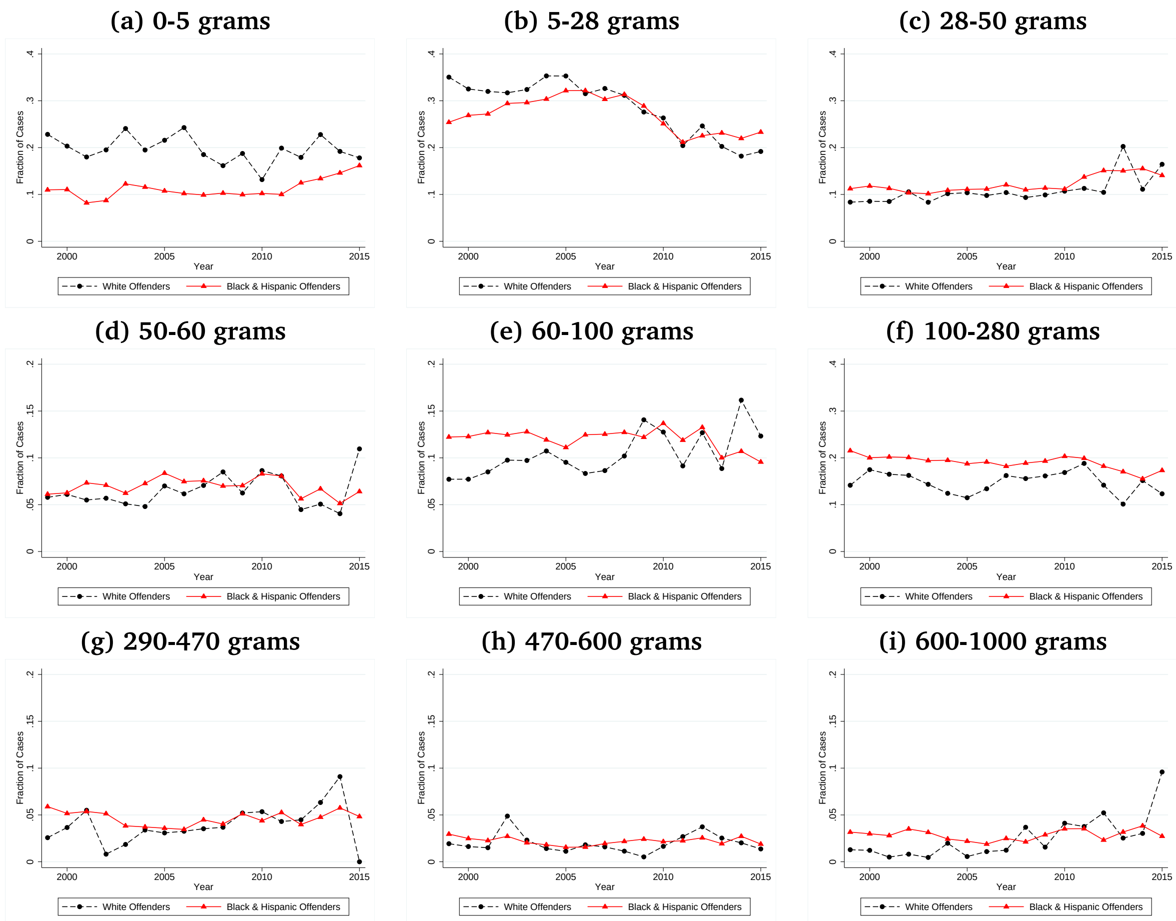
Notes. The figures above plot the scaled densities of drug quantities pre-2010 (in black) and the actual densities of drug quantities post-2010 (in red). The amounts are aggregated into 10-gram bins and limited to drug quantities under 1000g.

Figure 4. Post-2010 Density Minus Scaled Pre-2010 Density



Notes. The figure above plots the difference between the post-2010 density and the scaled density of drug quantities in pre-2010 for each 10-gram bin. Confidence intervals are calculated by bootstrapping as discussed in the text.

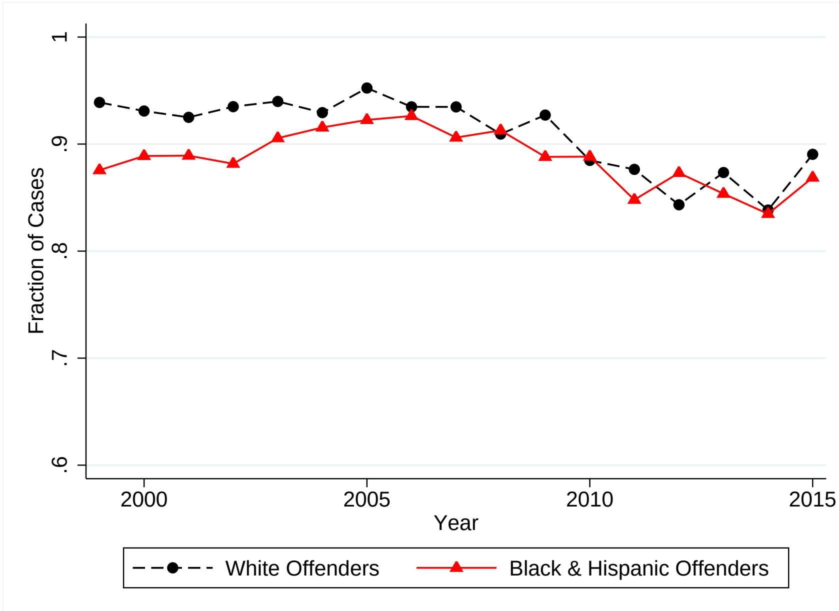
Figure 5. Changing Distribution of Drug Weights Over Time by Race.



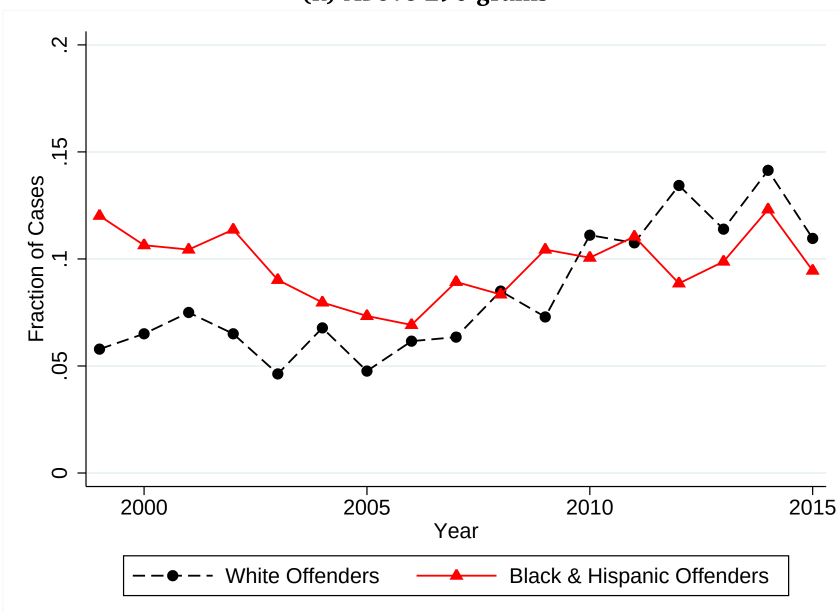
Notes. The figures above plot the share of cases in the specified range by year for white and black and Hispanic offenders. For example, panel (a) plots the share of cases with 0-5g (not including 5g) in each year from 1999-2015. Panel (b) plots the share of cases with 5-28g in each year from 1999-2015, and so on.

Figure 5. Changing Distribution of Drug Weights Over Time by Race.

(j) Below 280 grams



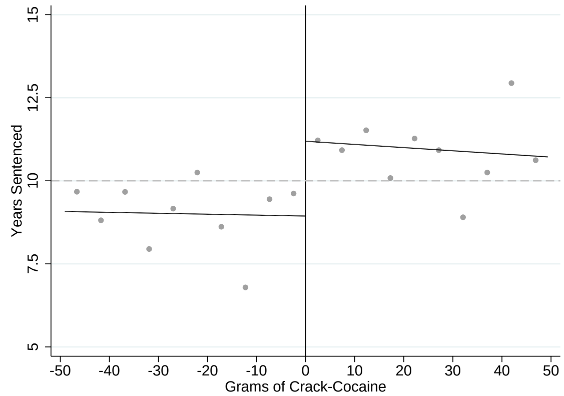
(k) Above 290 grams



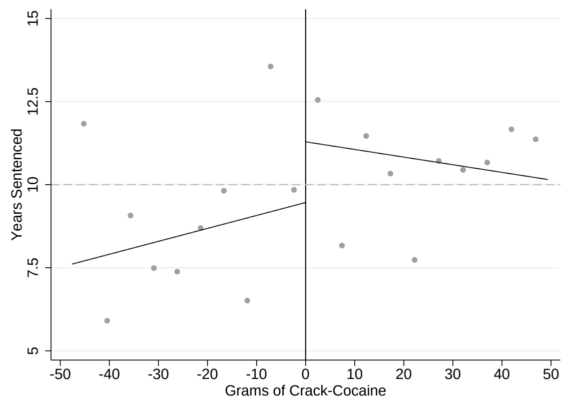
Notes. The figures above plot the share of cases in the specified range by year for white and black and Hispanic offenders. Panel (a) plots the share of cases below 280g (not including 280g) in each year from 1999-2015. Panel (b) plots the share of cases above 290g (including 290g) in each year from 1999-2015.

Figure 6. Sentencing Consequence of Crossing the Mandatory Minimum Threshold

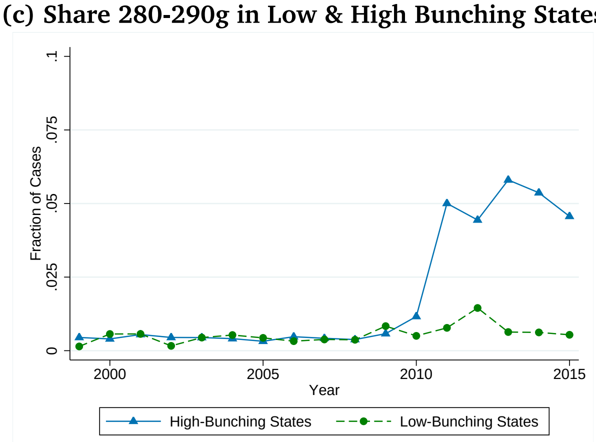
(a) All States



(b) Low-Bunching States

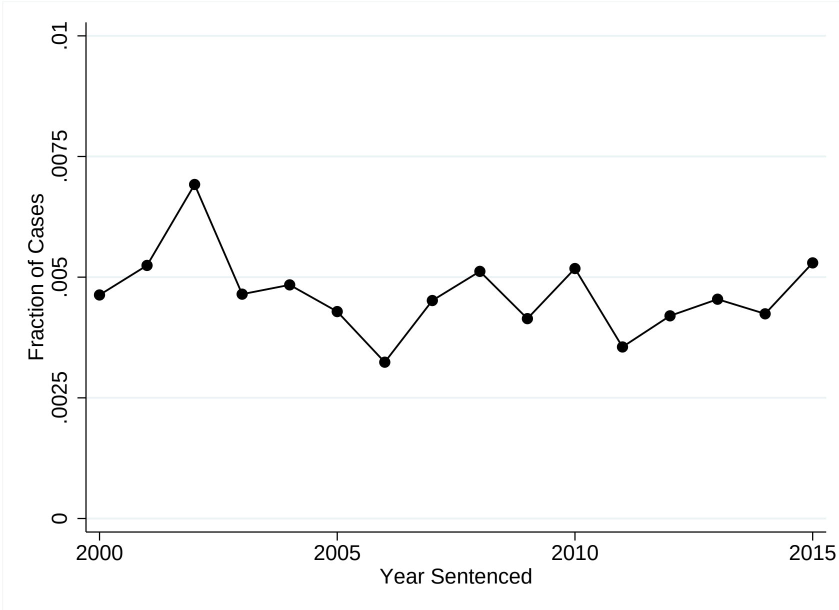


(c) Share 280-290g in Low & High Bunching States

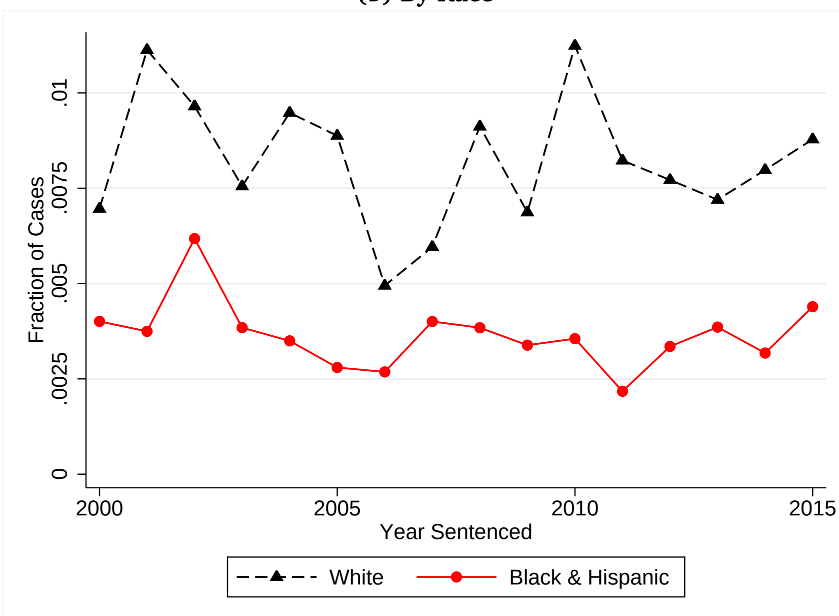


Notes. Figure 6a plots the average sentence (within each 5g bin) from 230-330g for cases sentenced after 2010. A linear fit is estimated on each side of the 280g threshold. The estimated sentencing discontinuity is about 2.25 years. Figure 6b is the same plot but limited to the subset of states that have low-levels of bunching. The estimated discontinuity is about 1.85 years. Figure 6c plots the share of cases with 280-290g by year for low- and high-bunching states.

Figure 7. Fraction of All Cocaine Offenses with 200-400g in FL
(a) All Offenders

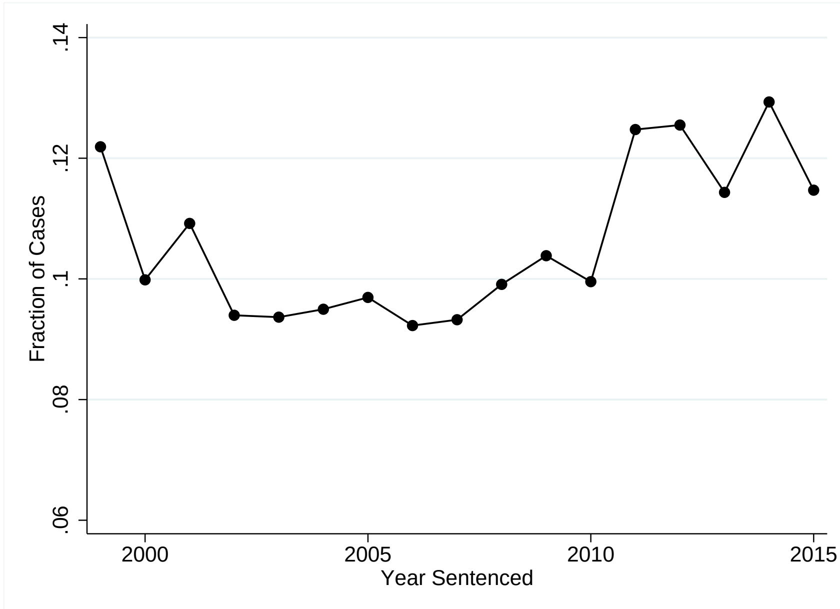


(b) By Race



Notes. The figures above plot the fraction of cocaine offenses that have a range from 200-400g in FL state prison from 1999-2015. The denominator is all cocaine offenses in FL. There is no break in the percent of offenses in those states after 2010.

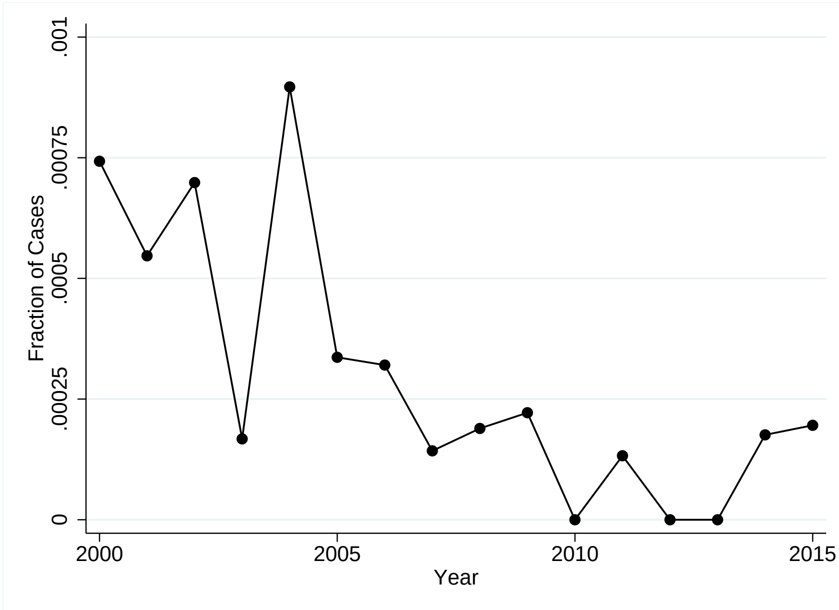
Figure 7. Percent of Cocaine Offenses from 200-400g
(c) USSC Federal Data



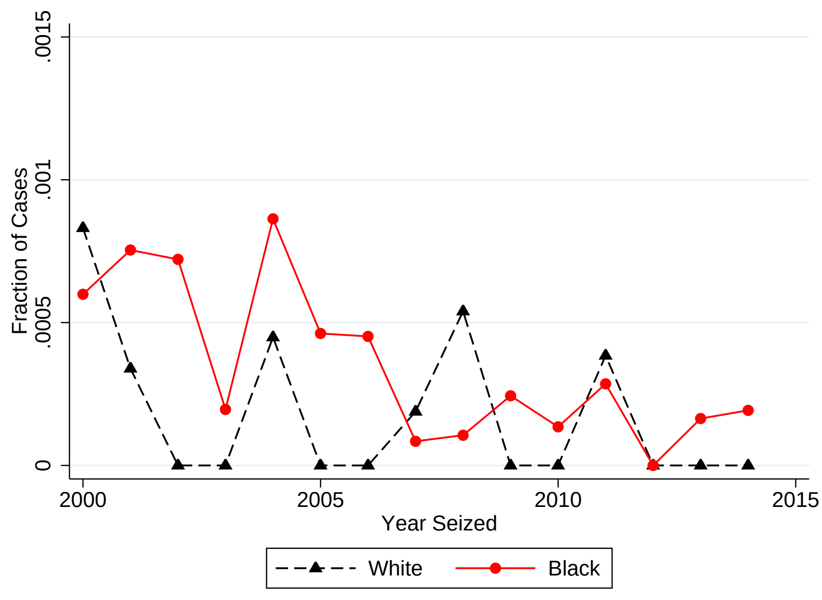
Notes. The figures above plot the fraction of crack-cocaine offenses that have a range from 200-400g in federal cases from 1999-2014. The denominator is all crack-cocaine offenses in the federal data. The fraction of cases in this broad range does exhibit a sharp increase after 2010.

Figure 8. Distribution of Drug Amounts Around 280g Pre- and Post-2010 from NIBRS.

(a) All Offenders

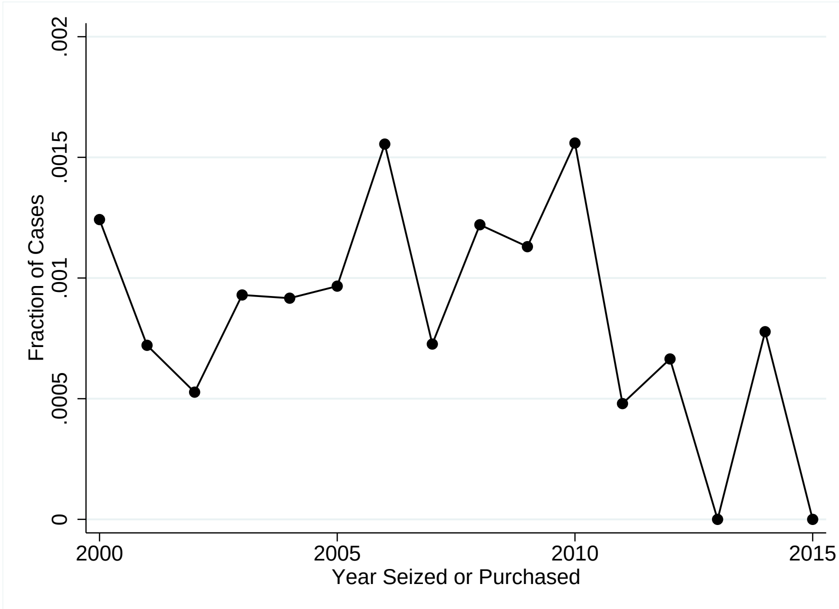


(b) By Race



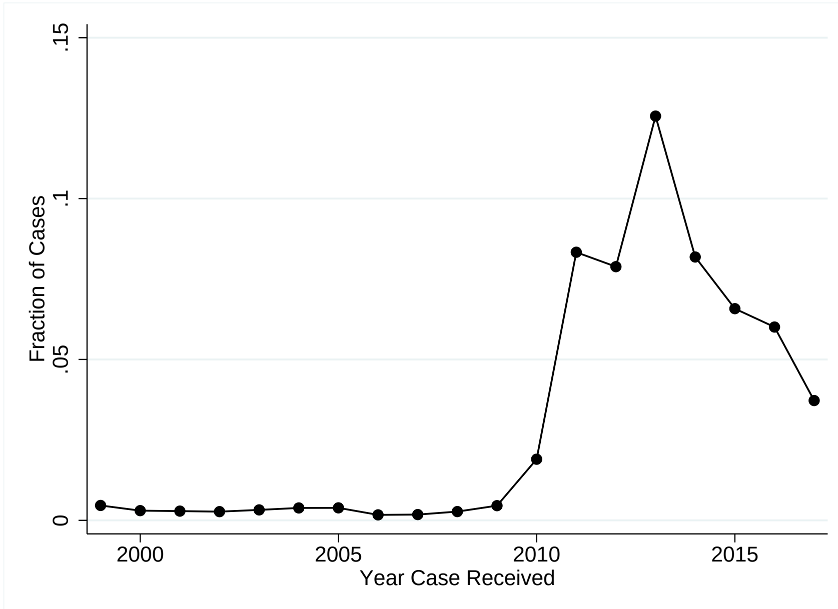
Notes. The figures above plot the distribution of drug amounts for crack-cocaine recorded from local police departments. Panel (a) displays this distribution over time for all offenders. Panel (b) displays this distribution over time by race.

Figure 9. Changing Fraction of Drug Exhibits with 280-290g in DEA STRIDE.



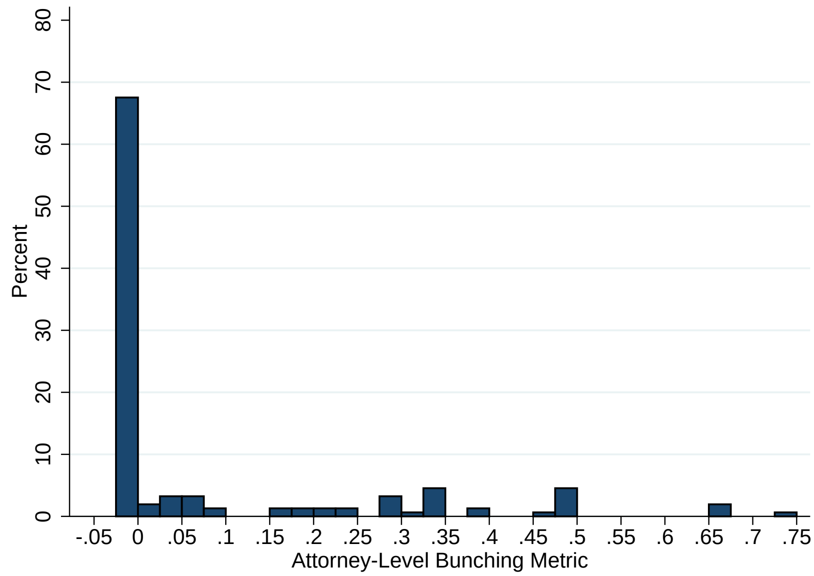
Notes. The figures above plot the fraction of cocaine drug exhibits sent to DEA laboratories and recorded as 280-290g from 2000-2015. The denominator is all cocaine exhibits in the DEA STRIDE data. Results are similar if limited to “cocaine hydrochloride” or “cocaine base.”

Figure 10. Changing Fraction of Cases with 280-290g in EOUSA System.



Notes. The figure above plots the fraction of crack-cocaine cases recorded as 280-290g in the EOUSA caseload data. The denominator is all crack-cocaine cases in the EOUSA data with non-missing drug quantities. The EOUSA data contains many more missing values than the USSC data. I believe this is an issue with differential record-keeping requirements by district. For example, in West Virginia North and Mississippi North districts, over 99% of cases are missing quantity data. Mississippi South and Alabama North, on the other hand, have quantity information for over 85% of cases.

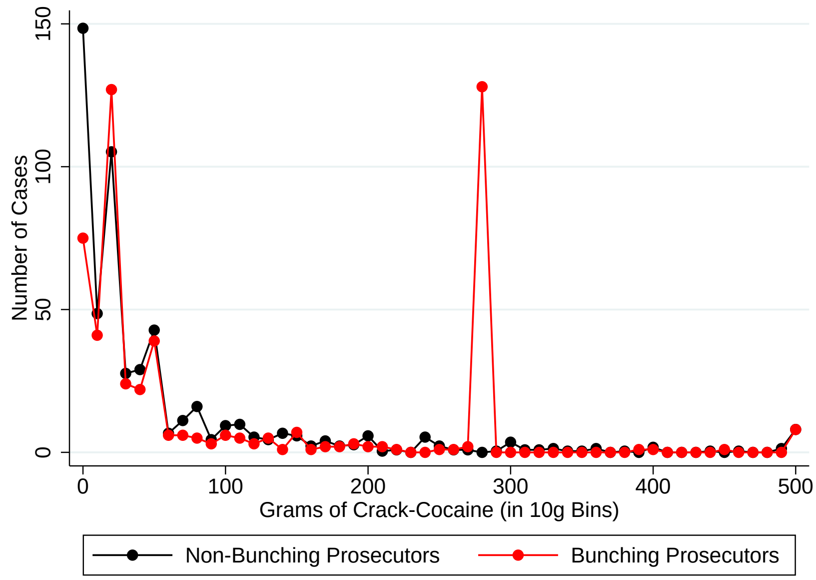
Figure 11. Distribution of Bunching at 280-290g after 2010 by Lead Attorney.



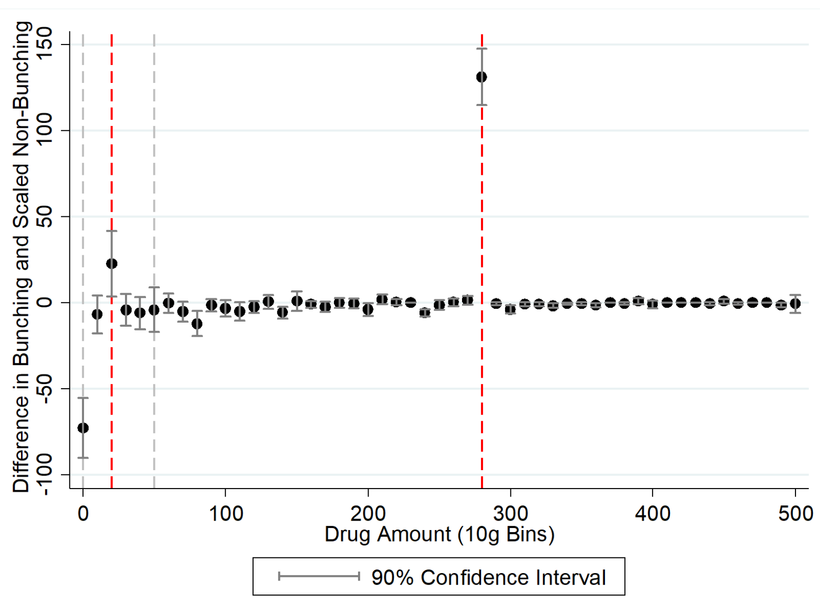
Notes. The figure above plots the distribution of the attorney-level bunching metric defined in equation (8) above. To calculate this metric, I first calculate the percentage of cases prior to 2010 that are recorded with 280-290g–about 0.4%. Then, for each attorney, I calculate the percentage of their cases post-2010 that are recorded with 280-290g. The denominator in that calculation is the total number of each attorney's cases with non-missing drug quantity and drug quantity less than 1000g. Finally, I subtract 0.4% from the attorney-level percentage. A "bunching metric" greater than zero implies that the attorney has a higher percentage of cases with 280-290g than normal. I only plot this metric for attorneys with at least 15 cases post-2010.

Figure 12. Comparison of Bunching and Non-Bunching Attorney Densities, Post-2010

(a) Densities of Crack-Cocaine Quantity

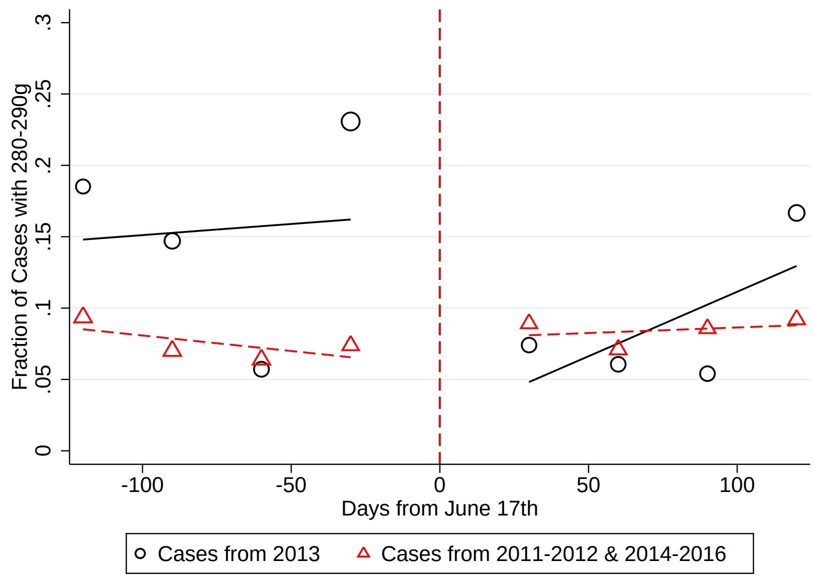


(b) Difference between Bunching Density and Non-Bunching Density



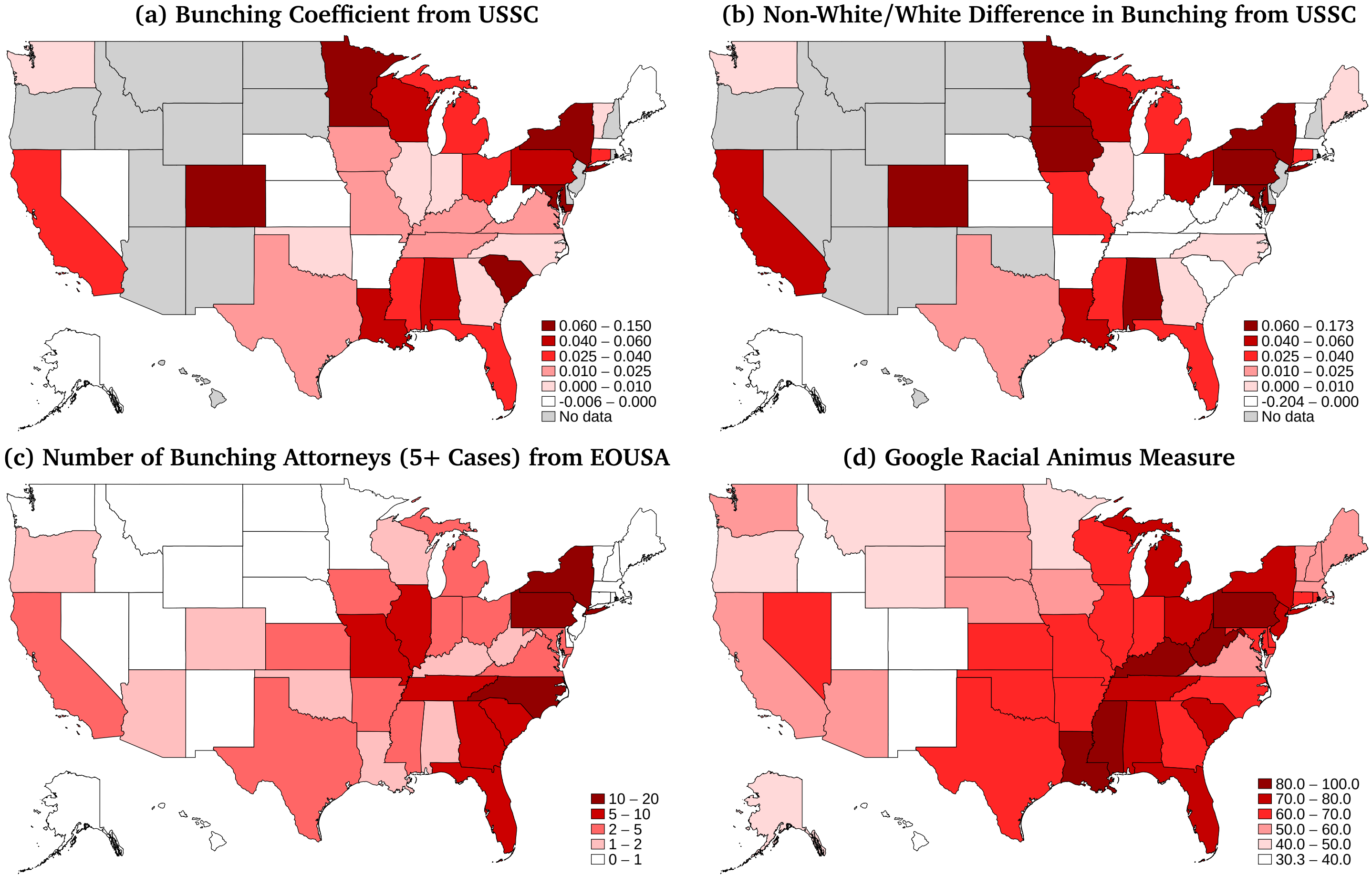
Notes. The figure in panel (a) plots the density of drug quantity for the attorneys who I identify as bunching attorney (bunching metric above zero) and those attorneys who I do not identify as bunching attorneys (bunching metric less than or equal to zero). The figure in panel (b) plots the difference between the bunching and non-bunching density for each 10g bin.

Figure 13. Change in Bunching by Prosecutors after *Alleyne* Decision



Notes. The scatter plot above shows the fraction of cases with 280-290g in each 30-day bin for 120 days before and 120 days after June 17th. The black circles show the fraction of cases in each bin for 2013 and the red triangles show the average fraction of cases in each bin for 2011-2012 and 2014-2016. The solid black line shows a linear fit on each side of the June 17, 2013 and the dashed red line shows a linear fit on each side of June 17 for all other years. The scatter plot symbols are weighted by the total number of cases in each bin.

Figure 14. Map of State-level Bunching and State-level Racial Disparity in Bunching



Notes. Panel (a) plots the state-level bunching estimate for all states with a sufficient number of cases. Panel (b) plots the difference between the state-level bunching estimate for white offenders and the state-level bunching estimate for black and Hispanic offenders for all states with a sufficient number of cases. Panel (c) plots the number of prosecutors who bunch in each state (among those prosecutors with 5+ drug cases after 2010). Panel (d) plots the racial animus index derived from Google search volume for a racial slur and introduced by Stephens-Davidowitz (2014).

Appendix A. Additional Analyses

Table A1. Result Robust to Sample Restrictions

	Pr(280-290g Crack-Cocaine)					
	(1)	(2)	(3)	(4)	(5)	(6)
After 2010	0.0135***			0.0133***		
	(0.00211)			(0.00216)		
After 2010 x White		0.00527	0.00523		0.00503	0.00520
		(0.00681)	(0.00681)		(0.00683)	(0.00683)
After 2010 x Non-White		0.0137***	0.0133***		0.0135***	0.0133***
		(0.00218)	(0.00225)		(0.00223)	(0.00230)
Constant	0.00436***	0.00439***	0.00443***	0.00462***	0.00463***	0.00446***
	(0.000363)	(0.000381)	(0.000397)	(0.000600)	(0.000602)	(0.000612)
P-value: W = NW	-	0.2362	0.2600	-	0.2363	0.2600
Only One Drug Charge	Yes	Yes	Yes	Yes	Yes	Yes
Post-2006 Data Only	No	No	No	Yes	Yes	Yes
Hispanic Offenders Excluded	No	No	Yes	No	No	Yes
Observations	37,113	34,147	31,760	16,844	16,810	15,642
R-squared	0.003	0.003	0.003	0.004	0.004	0.004

Notes. Robust standard errors in parentheses. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." The row "Only One Drug Charge" is equal to "Yes" when the sample includes offenders only charged with trafficking one type of drug. The row "Post-2006 Data Only" is equal to "Yes" when the data is limited to cases brought to court from 2006-2014. The row "Hispanic Offenders Excluded" is equal to "Yes" when Hispanic offenders are removed from the sample.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A2. Result Robust to Controls and Alternative Std. Errors.

	Pr(280-290g Crack-Cocaine)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After 2010	0.0349***		0.0351***		0.0348***		0.0330***	
	(0.00728)		(0.00726)		(0.00713)		(0.00639)	
After 2010 x White		0.0136*		0.0139*		0.0156**		0.0138*
		(0.00686)		(0.00692)		(0.00727)		(0.00694)
After 2010 x Non-White		0.0361***		0.0365***		0.0361***		0.0342***
		(0.00759)		(0.00765)		(0.00751)		(0.00679)
Constant	0.00515***	0.00520***	0.00891***	0.00893***	0.00945***	0.00944***	0.00838**	0.00836**
	(0.000452)	(0.000496)	(0.00246)	(0.00247)	(0.00260)	(0.00261)	(0.00319)	(0.00320)
P-value: W = NW	-	0.0243	-	0.0227	-	0.0439	-	0.0445
Offender Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
Year Trend	No	No	No	No	No	No	Yes	Yes
Observations	55,469	51,336	50,462	50,396	50,454	50,396	50,454	50,388
R-squared	0.015	0.016	0.016	0.016	0.020	0.021	0.019	0.021

Notes. Standard errors clustered at the state-level in parentheses. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." The row "Offender Controls" indicates if the following offender-level controls are included: criminal history points, age, citizenship, number of current offense counts, whether a weapon was involved, and education. The rows "State Fixed Effects" and "Year Trend" indicate if the specification includes state fixed effects or a year trend as controls.
*** p<0.01, ** p<0.05, * p<0.1

Table A3. Result Robust to Probit & Logit Models.

	Pr(280-290g Crack-Cocaine)			
	Probit		Logit	
	(1)	(2)	(3)	(4)
After 2010	0.816***		2.089***	
	(0.0336)		(0.0857)	
After 2010 x White		0.484***		1.301***
		(0.137)		(0.343)
After 2010 x Non-White		0.827***		2.112***
		(0.0346)		(0.0883)
Constant	-2.567***	-2.564***	-5.268***	-5.258***
	(0.0221)	(0.0231)	(0.0642)	(0.0668)
P-value: W = NW	-	0.0124	-	0.0791
Observations	55,694	51,535	55,694	51,535

Notes. Robust standard errors in parentheses. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White."
*** p<0.01, ** p<0.05, * p<0.1

Table A4. Result Robust to Other Categorizations of Bunching

	Pr(280-300g)	Pr(280-320g)	Pr(280-380g)
	(1)	(2)	(3)
After 2010 x White	0.0170**	0.0135*	0.0124
	(0.00714)	(0.00770)	(0.00937)
After 2010 x Non-White	0.0364***	0.0373***	0.0399***
	(0.00240)	(0.00264)	(0.00307)
Constant	0.00803***	0.0156***	0.0314***
	(0.000428)	(0.000595)	(0.000836)
P-value: W = NW	0.0098	0.0033	0.0049
Observations	51,535	51,535	51,535
R-squared	0.013	0.008	0.006

Notes. Robust standard errors in parentheses. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White."
*** p<0.01, ** p<0.05, * p<0.1

Table A5. Result Robust to Other Sample Restrictions

	Pr(280-290g Crack-Cocaine)				
	(1)	(2)	(3)	(4)	(5)
After 2010 x White	0.0128**	0.0124**	0.0124**	0.00476***	0.0997***
	(0.00587)	(0.00566)	(0.00566)	(0.00227)	(0.0164)
After 2010 x Non-White	0.0347***	0.0332***	0.0330***	0.00912***	0.1305***
	(0.00222)	(0.00212)	(0.00211)	(0.000909)	(0.0048)
Constant	0.00490***	0.00471***	0.00466***	-0.00134***	0.0058***
	(0.000326)	(0.000313)	(0.000310)	(0.000132)	(0.00033)
Sample Restriction	0-2500g	0-25000g	No Restriction	0-1000g	0-1000g
Bandwidth Dummy (270-290g)	No	No	No	Yes	No
Includes Weights Coded as a Range	No	No	No	No	Yes
P-value: W = NW	0.0005	0.0006	0.0006	0.0679	0.0709
Observations	54,303	56,592	57,099	51,535	58,036
R-squared	0.015	0.014	0.014	0.737	0.081

Notes. Robust standard errors in parentheses. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White." Column (4) reports results from a regression that includes a dummy variable equal to one if the drug weight is 270-290g. This limits the identifying variation to those offenses from 270-290g. Column (5) reports results when the sample includes quantities coded as a range (in this analysis, the lower bound of the range is used).
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A6. Missing Mass in the Distribution of Drug Amounts, with Linear Time Trend and State FEs (No Interactions)

Panel A. Analysis of Changes in the 0-100g Range.

	Pr(0-5g)	Pr(5-28g)	Pr(28-50g)	Pr(50-60g)	Pr(60-100g)
	(1)	(2)	(3)	(4)	(5)
After 2010 x White	0.0107	-0.1249***	0.0291*	-0.0119	0.0172
	(0.0180)	(0.0198)	(0.0155)	(0.0114)	(0.0150)
After 2010 x Non-White	0.0237***	-0.0848***	0.0318***	-0.0165***	-0.0142***
	(0.0050)	(0.0069)	(0.0053)	(0.0041)	(0.0052)
Constant	0.1700***	0.3269***	0.1026***	0.0782***	0.1096***
	(0.0070)	(0.0089)	(0.0059)	(0.0049)	(0.0060)
P-value: W = NW	0.4740	0.0430	0.8635	0.6863	0.0370
Observations	52,530	52,530	52,530	52,530	52,530
R-squared	0.031	0.019	0.006	0.007	0.006

Panel B. Analysis of Changes in the 100-1000g Range.

	Pr(100-280g)	Pr(280-290g)	Pr(290-470g)	Pr(470-600g)	Pr(600-1000g)
	(1)	(2)	(3)	(4)	(5)
After 2010 x White	0.0125	0.0120**	0.0175*	0.0117*	0.0260***
	(0.0169)	(0.0055)	(0.0098)	(0.0071)	(0.0088)
After 2010 x Non-White	-0.0005	0.0344***	0.0128***	0.0067***	0.0067**
	(0.0062)	(0.0023)	(0.0034)	(0.0024)	(0.0027)
Constant	0.1499***	0.0059***	0.0291***	0.0132***	0.0146***
	(0.0071)	(0.0013)	(0.0038)	(0.0026)	(0.0027)
P-value: W = NW	0.4414	0.0001	0.6314	0.4868	0.0279
Observations	52,530	52,530	52,530	52,530	52,530
R-squared	0.012	0.023	0.007	0.006	0.010

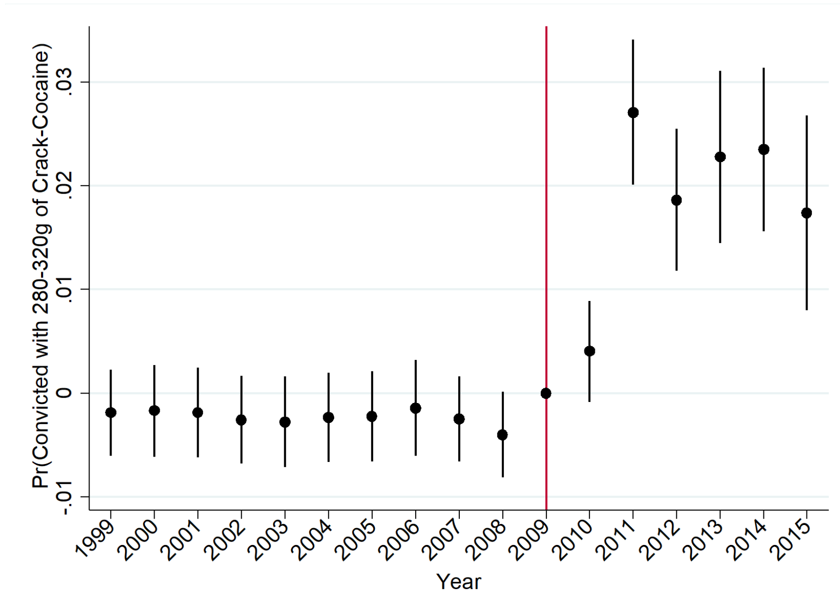
Notes. Robust standard errors in parentheses. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams, and all specifications include state fixed effects and a linear time trend. In addition, specifications in this table include an interaction between the After 2010 dummy and the linear time trend. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White."
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A7. Missing Mass in the Distribution of Drug Amounts, with Quadratic Time Trend Interaction and State FEs

Panel A. Analysis of Changes in the 0-100g Range.					
	Pr(0-5g)	Pr(5-28g)	Pr(28-50g)	Pr(50-60g)	Pr(60-100g)
	(1)	(2)	(3)	(4)	(5)
After 2010 x White	0.0528	-0.0269	0.0147	-0.0129	-0.0259
	(0.0360)	(0.0414)	(0.0301)	(0.0254)	(0.0298)
After 2010 x Non-White	0.0050	-0.0358***	0.0203**	0.0036	-0.0153*
	(0.0077)	(0.0109)	(0.0085)	(0.0069)	(0.0085)
Constant	0.1215***	0.2362***	0.0994***	0.0975***	0.1315***
	(0.0236)	(0.0296)	(0.0200)	(0.0174)	(0.0213)
P-value: W = NW	0.1943	0.8347	0.8575	0.5313	0.7334
Observations	52,530	52,530	52,530	52,530	52,530
R-squared	0.032	0.021	0.006	0.008	0.007
Panel B. Analysis of Changes in the 100-1000g Range.					
	Pr(100-280g)	Pr(280-290g)	Pr(290-470g)	Pr(470-600g)	Pr(600-1000g)
	(1)	(2)	(3)	(4)	(5)
After 2010 x White	0.0010	0.0133	-0.0280	0.0169	-0.0050
	(0.0367)	(0.0099)	(0.0202)	(0.0139)	(0.0175)
After 2010 x Non-White	-0.0036	0.0302***	-0.0018	-0.0028	0.0004
	(0.0102)	(0.0040)	(0.0055)	(0.0038)	(0.0046)
Constant	0.1926***	0.0038	0.0602***	0.0095	0.0478***
	(0.0240)	(0.0040)	(0.0139)	(0.0080)	(0.0112)
P-value: W = NW	0.9032	0.1141	0.2108	0.1717	0.7655
Observations	52,530	52,530	52,530	52,530	52,530
R-squared	0.012	0.023	0.008	0.006	0.010

Notes. Robust standard errors in parentheses. All specifications above use the sample of offenses with drug amounts between 0 grams and 1000 grams, and all specifications include state fixed effects and a linear time trend. In addition, specifications in this table include an interaction between the After 2010 dummy and the linear time trend. The row "P-value: W = NW" reports the p-value from a test of the null hypothesis that the coefficient on "After 2010 x White" is equal to the coefficient on "After 2010 x Non-White."
*** p<0.01, ** p<0.05, * p<0.1

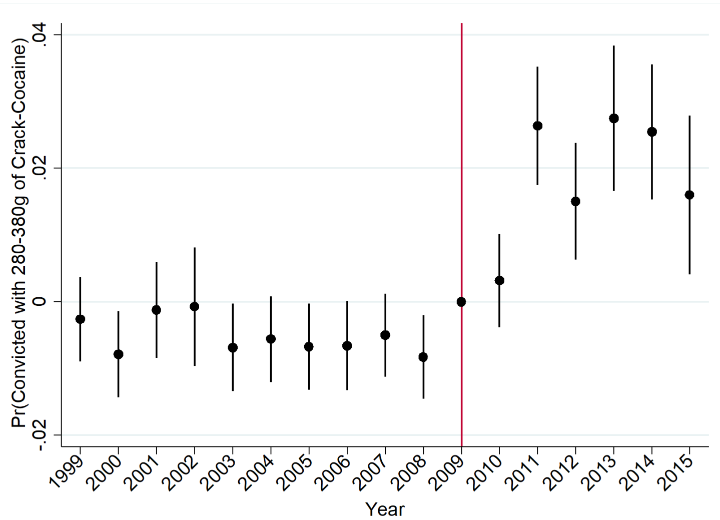
Figure A1a. Bunching Coefficient by Year.



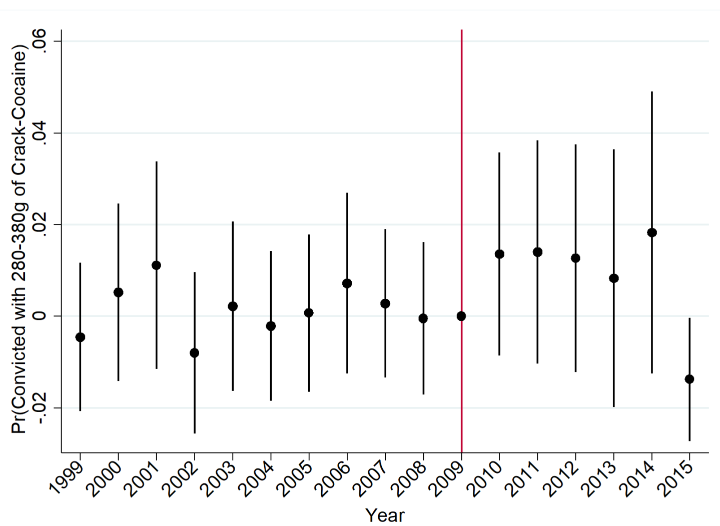
Notes. I estimate the main bunching coefficient by year (relative to 2010) and plot the coefficients with 90% confidence intervals in the figure above. Points to the left of the red line are prior to 2010 and points to the right of the red line are after 2010. The dependent variable used is Pr(Holding 280-320g) instead of Pr(Holding 280-290g). The broader range gives me the statistical power necessary to estimate the coefficient by year.

Figure A1. Bunching Coefficient by Year and Race

(b) Black and Hispanic Offenders

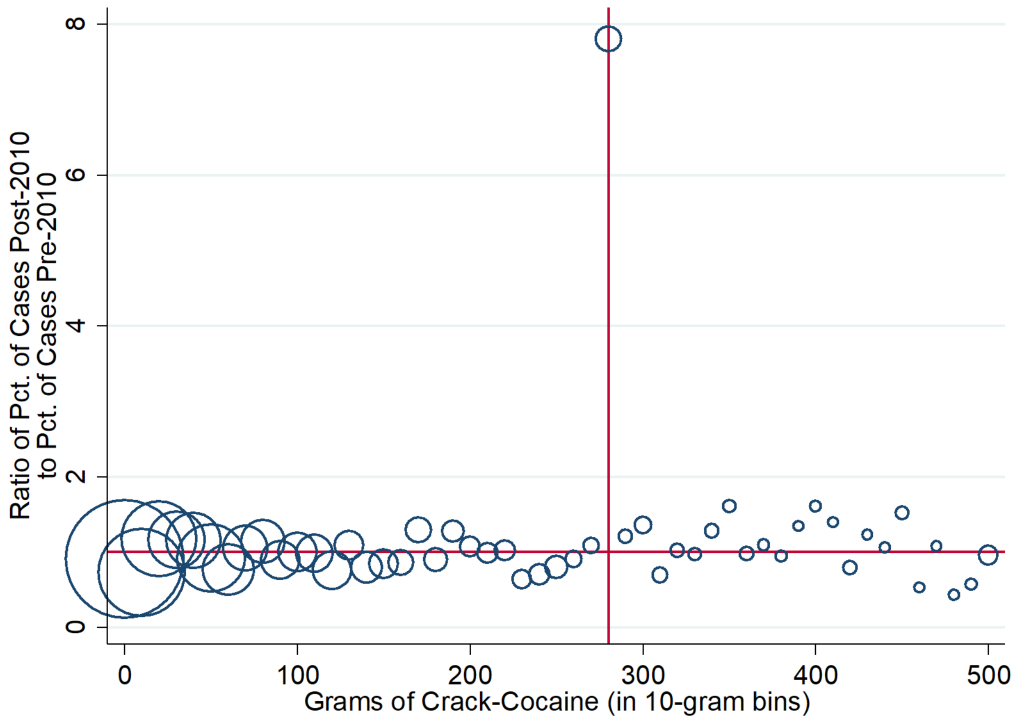


(c) White Offenders



Notes. I estimate the main bunching coefficient by race and year (relative to 2010) and plot the coefficients with 90% confidence intervals in the figure above. Points to the left of the red line are prior to 2010 and points to the right of the red line are after 2010. Panel (a) displays this plot for black and Hispanic offenders and Panel (b) displays this plot for white offenders. The dependent variable used is Pr(Holding 280-320g) instead of Pr(Holding 280-290g). The broader range gives me the statistical power necessary to estimate the coefficient by year.

Figure A2a. Bunching Ratio from 0-500g.



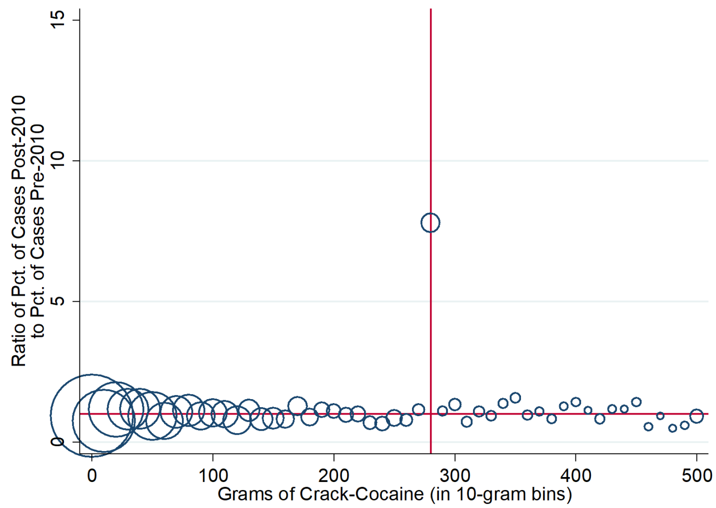
Notes. The figure above plots the bunching ratio for each 10-gram bin from 0-500 grams. The bunching ratio for each bin b is defined as follows:

$$\text{Bunching Ratio}_b = \frac{\text{\% of cases in b post-2010}}{\text{\% of cases in b pre-2010}}$$

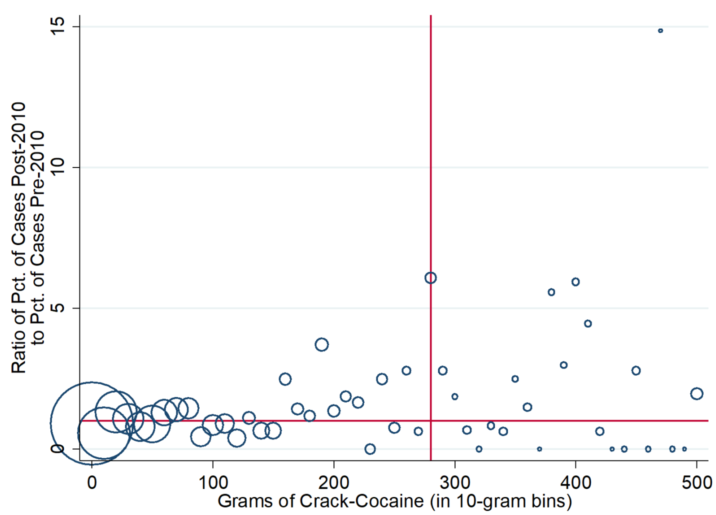
If the distributions were the same pre- and post-2010, the bunching ratio will equal 1 (marked by the horizontal red line). If the ratio is above 1, there is a higher degree of bunching in bin b post-2010. If the ratio is below 1, there is a lower degree of bunching post-2010. Each bin b is weighted by the total number of cases in the bin pre- and post-2010.

Figure A2. Bunching Ratio from 0-500g by Race

(b) Black and Hispanic Offenders



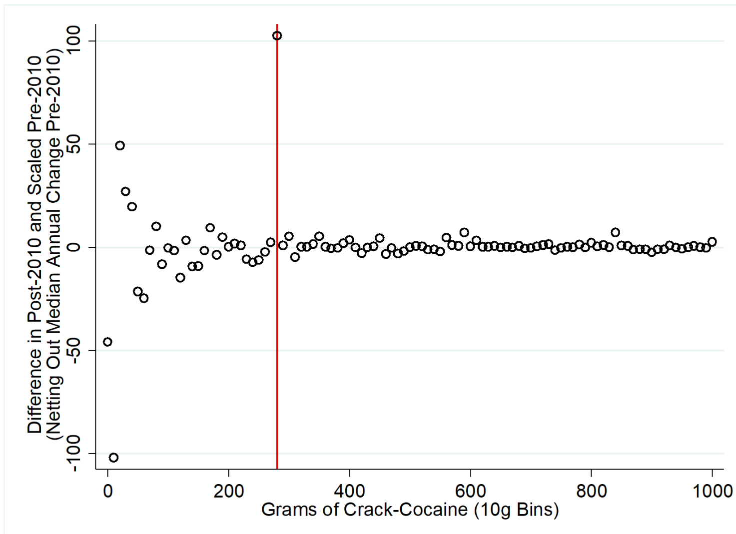
(c) White Offenders



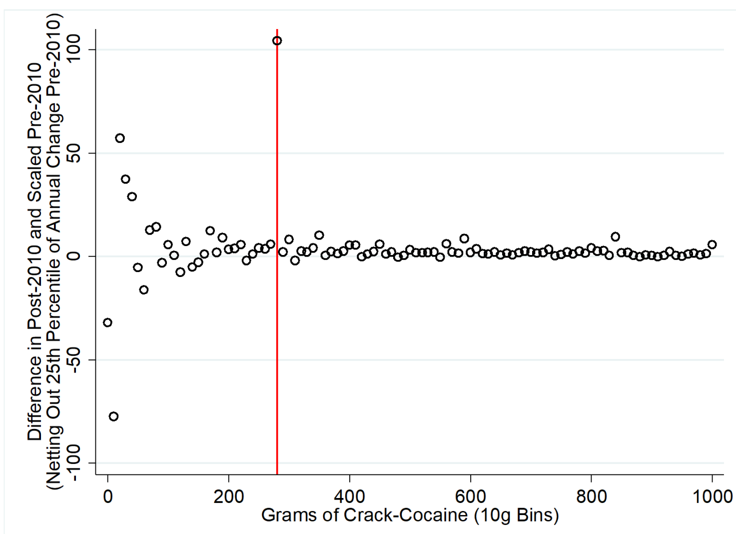
Notes. The figures above plot the bunching ratio for each 10-gram bin from 0-500 grams for black and Hispanic and white offenders. If the distributions were the same pre- and post-2010, the bunching ratio will equal 1 (marked by the horizontal red line). If the ratio is above 1, there is a higher degree of bunching in bin b post-2010. If the ratio is below 1, there is a lower degree of bunching post-2010. Panel (a) plots this ratio for black and Hispanic offenders and Panel (b) plots this ratio for white offenders. Each bin b is weighted by the total number of cases in the bin pre- and post-2010.

Figure A3. Post-2010 Density Minus Pre-2010 Density, Netting Out Bin Variation in Pre-2010 Years

(a) Netting out Median Annual Difference Pre-2010



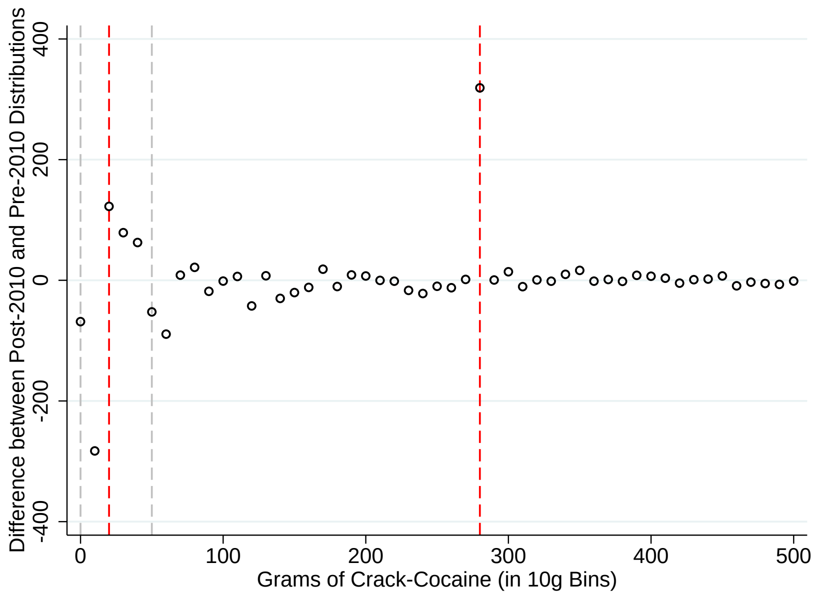
(b) Netting out 25th Percentile of Annual Difference Pre-2010



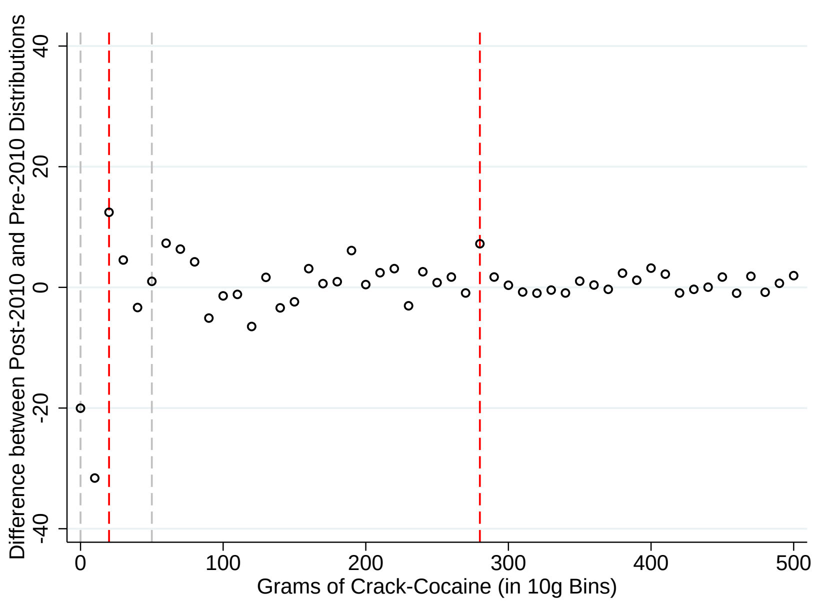
Notes. These figures plot the difference (averaged by year) between the post-2010 density and the scaled pre-2010 density for each 10-gram bin, netting out the median and 25th percentile of annual differences from 1999-2009. To calculate the difference between post-2010 and scaled pre-2010, I scale all years to have the same number of cases as the year 2011. Then, I average the years 2011-2014 and the years 1999-2009 and take the difference in those averages for each 10-gram bin. To calculate the annual change from 1999-2009, I substract the year 1999's density from the year 2000's, the year 2000's from the year 2001's, etc. I then take the median and the 25th percentile of those differences and subtract those statistics (for each 10-gram bin) from the average pre- and post- difference to arrive at the figures above.

Figure A4. Post-2010 Density Minus Pre-2010 Density, By Race

(a) Black and Hispanic Offenders



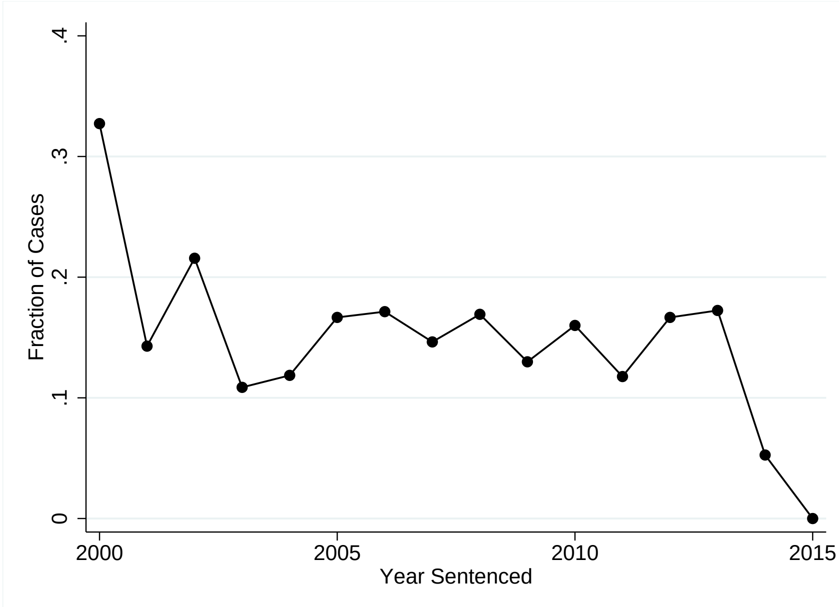
(b) White Offenders



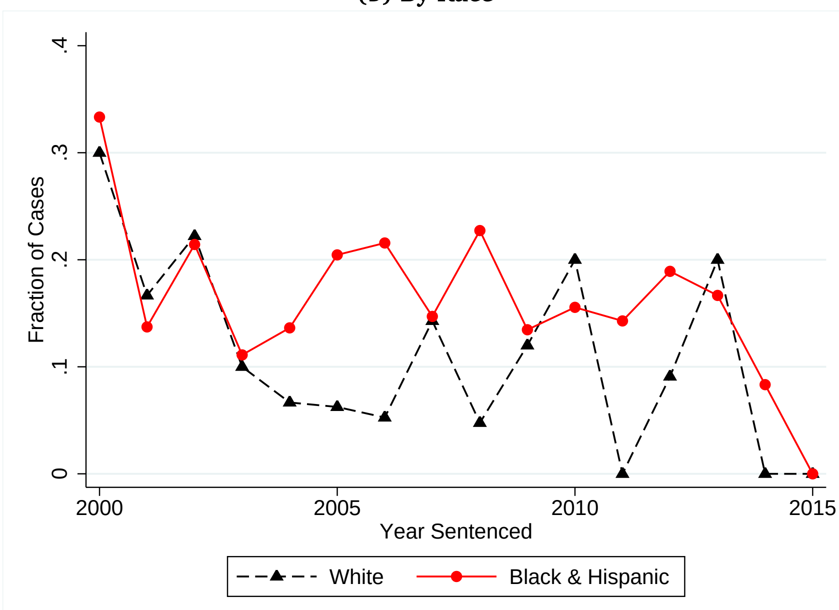
Notes. The figures above plot the difference between the post-2010 density and the scaled density of drug quantities in pre-2010 for each 10-gram bin by race.

Figure 5. Fraction of All Cocaine Offenses with 200-400g in NC

(a) All Offenders

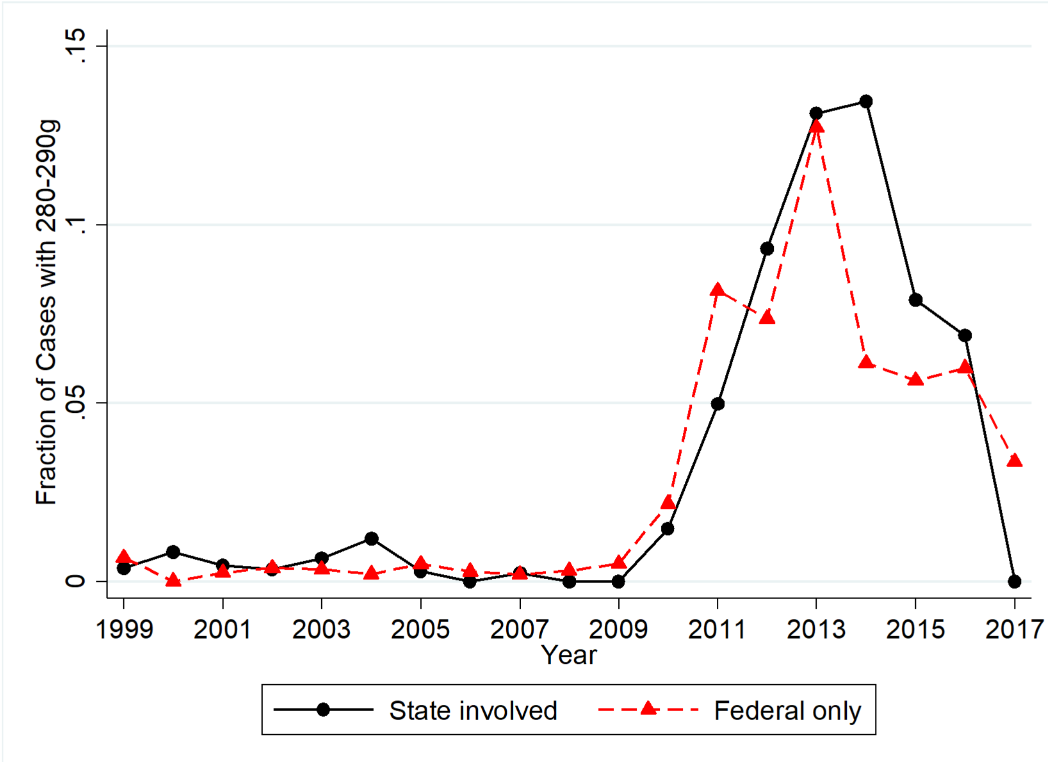


(b) By Race



Notes. The figures above plot the fraction of cocaine offenses that have a range from 200-400g in NC state prison from 2000-2015. The denominator is all cocaine offenses in NC. There is no break in the percent of offenses in those states after 2010.

Figure A6. Total Number of Cocaine Offenses in FL & NC



Notes. The figures above plot the total number of offenses classified as cocaine offenses by race in FL, MO, MN, NC, and WA state prison and sentencing data from 2001-2015. I overlay a cubic polynomial estimated on each side of 2010. There is no break in the number of offenses in those states after 2010.

Appendix B. Alternative Bunching Estimation

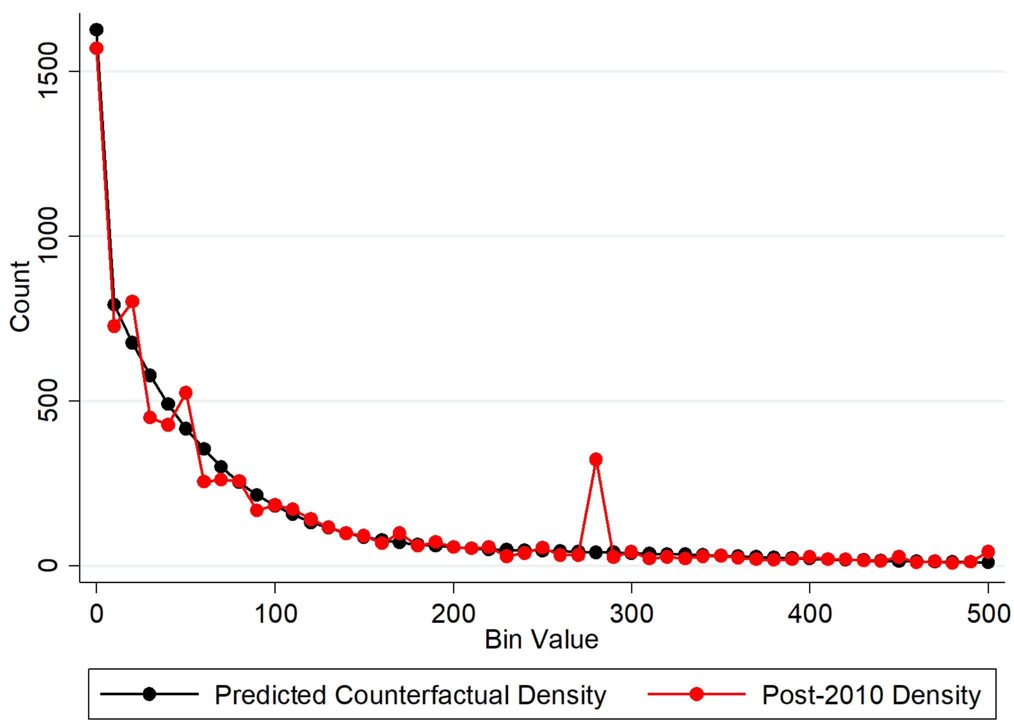
In order to estimate the degree of bunching at a given point in a density (of drug amounts or property crime values), I first construct a counterfactual density, an estimation of what the density might look like in the absence of bunching. One popular approach pioneered by Saez (2010) estimates the counterfactual density by using the actual bunched density. The main results in my paper rely on a different approach–constructing the counterfactual density from years where the notches are different. Kleven (2016) describes that approach as a "difference-in-bunching" design. Below, I apply the Saez (2010) method and show that it yields the same results.

First, I construct the counterfactual density by aggregating the data to 10-gram bins, summing the number of cases in each bin. With this aggregated data, I estimate a regression of the bin counts on a seventh-order polynomial of the bin values, dummies for the 270g and 280g bins, and a dummy for the 0g bin.

$$Count_b = \alpha_0 + \sum_{i=1}^{7} \beta_i (Amount_b)^i + \gamma_1 Bin270_b + \gamma_2 Bin280_b + \delta_1 Bin0_b + \epsilon_b \tag{1}$$

where $Count_b$ is the total number of cases in bin b, $Amount_b$ is the value of bin b, and $Bin[X]_b$ is a dummy variable indicating if the bin's value equals X. I use the parameter estimates from (8) (excluding γ_1 and γ_2) to predict a smooth density of bin counts. Furthermore, I adjust the predicted counts to force the smooth density to have the same number of cases as the actual density. I plot the counterfactual density and the actual post-2010 density below.

Figure B1. Predicted Counterfactual Density and Post-2010 Density



Notes. In the figure above, I plot a predicted counterfactual density of drug quantities (in black) and the actual density of drug quantities post-2010 (in red). The amounts are aggregated into 10-gram bins and limited to drug quantities under 500g.

Using the predicted counts from the counterfactual density and the actual counts post-2010, I construct the percent of cases in each bin for each density. I then calculate the difference in these percentages and run the following regression, bootstrapping the standard errors from 200 replications:

$$(\text{\% in Post2010 - \% in Predicted})_b = \alpha + \beta Bin280_b + \epsilon_b$$

The resulting $\beta = 0.0352$ and $SE_\beta = 0.0169$.

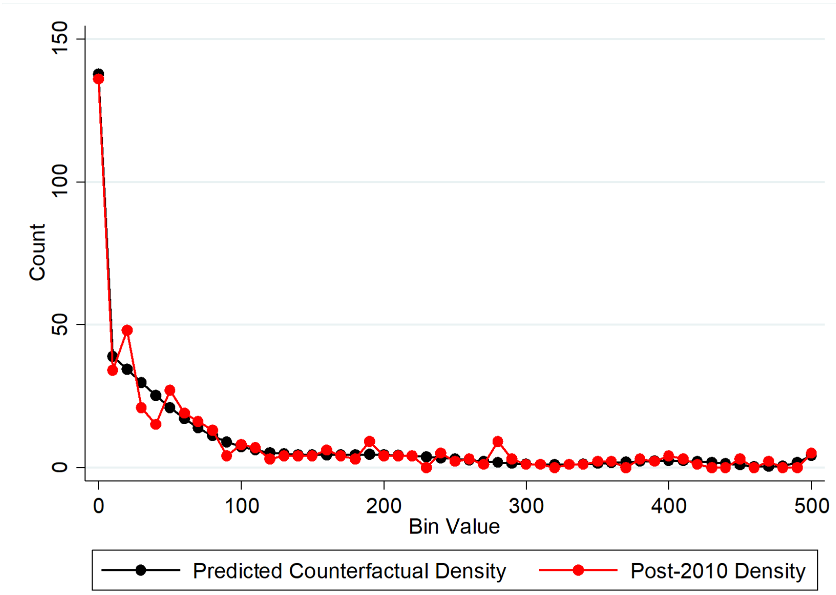
Next, I estimate:

$$(\text{\% in Post2010 - \% in Counterfactual})_{br} = \alpha + \beta Bin280_b + \gamma NonWhite_r + \delta Bin280_b \times NonWhite_r + \epsilon_b$$

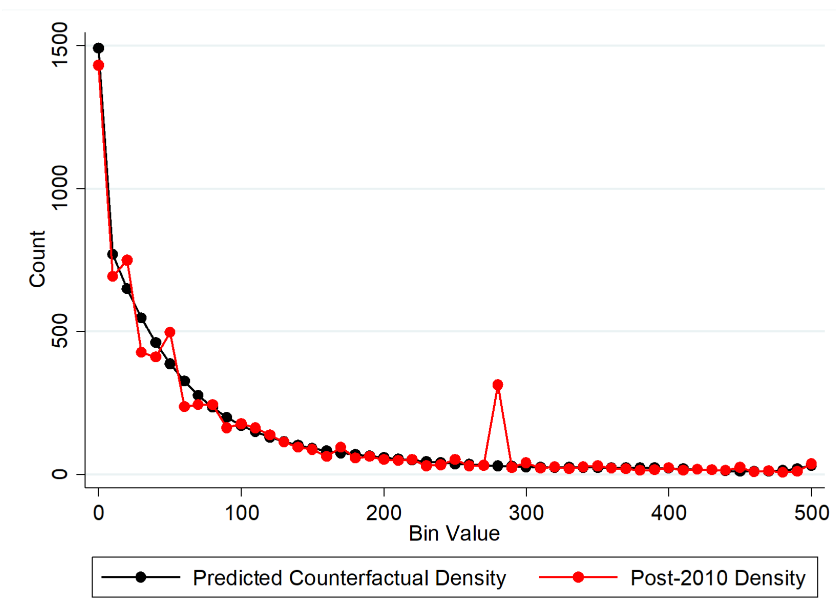
Using the Saez (2010) method, I estimate $\delta = 0.0237$ and $SE_\delta = 0.0119$. Using the difference-in-bunching method, I estimate $\delta = 0.0216$ and $SE_\delta = 0.0109$. In all analyses, I detect substantial bunching after 2010 and disproportionate bunching after 2010 for black and Hispanic offenders.

Figure B2. Pre-2010 Density and Post-2010 Density

(a) White Offenders, Saez (2010) Method



(b) Black and Hispanic Offenders, Saez (2010) Method



Notes. In the figure above, I plot predicted counterfactual densities of drug quantities (in black) and the actual densities of drug quantities post-2010 (in red) by race. The amounts are aggregated into 10-gram bins and limited to drug quantities under 500g.

In the method described above, I use case-level data to estimate bunching after 2010. However, I also compare aggregated pre-2010 and post-2010 data. To do this, I first scale the pre-2010 density to have the same total number of cases as the post-2010 density. Using the scaled counts from the counterfactual (pre-2010) density and the actual counts post-2010, I construct the percent of cases in each bin for each density. I then calculate the difference in these percentages and run the following regression, bootstrapping the standard errors from 200 replications:

$$(\text{\% in Post2010 - \% in Pre2010})_b = \alpha + \beta Bin280_b + \epsilon_b \tag{2}$$

where $Bin280_b$ is equal to one when bin b is the 280-290g bin and is equal to zero for all other bins.

Next, I re-construct the densities by race, scaling the pre-2010 densities separately by race to form two different counterfactual densities. I estimate:

$$(\text{\% in Post2010 - \% in Counterfactual})_{br} = \alpha + \beta Bin280_b + \gamma NonWhite_r + \delta Bin280_b \times NonWhite_r + \epsilon_{br} \tag{3}$$

Using the aggregated data, I also compare the difference between the actual and counterfactual density for every 10-gram bin to highlight bins with "missing mass" after 2010.[1] I consider the summed negative changes below 280g and the summed negative changes above 290g as measures of the total potential missing mass on each side of the threshold. For visual evidence, I also plot these differences for every 10-gram bin from 0 to 1000g.

[1] Since this exercise narrows the data to 10-gram bins, I do not use state fixed effects or time trends.

Appendix C. Conceptual Model of Offender Actions

Defendant chooses **p** (amount of drugs to carry) to solve the following:

$$\max_p b(p) - c(p) - s^0(p)$$

where $b(p)$ is the benefit associated with a given level of **p**, $c(p)$ is the cost associated with a given level of **p**, and $s^0(p)$ is the minimum sentence associated with a given level of **p**. Both benefits and costs are increasing in **p**, i.e. $b'(p) > 0$ and $c'(p) > 0$. The sentence schedule is:

$$s^0(p) = \begin{cases} 1 & p < t_L^0 \\ 5 & t_H^0 > p \geq t_L^0 \\ 10 & p \geq t_H^0 \end{cases}$$

Ignoring $s(p)$, $\mathbf{p}^*$ solves $b'(p) = c'(p)$. Defendant evaluates $\mathbf{p}^*$ against alternatives $t_L^0 - \varepsilon$ and $t_H^0 - \varepsilon$, where ε is an arbitrarily small amount.
If $p* < t_L^0$, then: defendant chooses $p^c = p*$.
If $t_H^0 > p* \geq t_L^0$ then:

- defendant chooses:
  - $p^c = p*$ if $b(p*) - c(p*) - 5 \geq b(t_L^0 - \varepsilon) - c(t_L^0 - \varepsilon) - 1$
- defendant chooses:
  - $p^c = t_L^0 - \varepsilon$ if $b(p*) - c(p*) - 5 < b(t_L^0 - \varepsilon) - c(t_L^0 - \varepsilon) - 1$

If $p* \geq t_H^0$ then:

- defendant chooses:
  - $p^c = p*$ if $b(p*) - c(p*) - 10 \geq b(t_H^0 - \varepsilon) - c(t_H^0 - \varepsilon) - 5$
  - and $b(p*) - c(p*) - 10 \geq b(t_L^0 - \varepsilon) - c(t_L^0 - \varepsilon) - 1$
- defendant chooses:
  - $p^c = t_H^0 - \varepsilon$ if $b(p*) - c(p*) - 10 < b(t_H^0 - \varepsilon) - c(t_H^0 - \varepsilon) - 1$
  - and $b(t_L^0 - \varepsilon) - c(t_L^0 - \varepsilon) - 5 < b(t_H^0 - \varepsilon) - c(t_H^0 - \varepsilon) - 1$
- defendant chooses:
  - $p^c = t_L^0 - \varepsilon$ if $b(p*) - c(p*) - 10 < b(t_L^0 - \varepsilon) - c(t_L^0 - \varepsilon) - 1$
  - and $b(t_H^0 - \varepsilon) - c(t_H^0 - \varepsilon) - 5 < b(t_L^0 - \varepsilon) - c(t_L^0 - \varepsilon) - 1$

Suppose sentencing schedule changes to the following:

$$s^1(p) = \begin{cases} 1 & p < t_L^1 \\ 5 & t_H^1 > p \geq t_L^1 \\ 10 & p \geq t_H^1 \end{cases}$$

where $t_L^1 > t_L^0$ and $t_H^1 > t_H^0$ and $t_H^0 > t_L^1$. Now...
If $p* < t_L^0$, then: defendant chooses $p^c = p*$.
If $t_L^1 > p* \geq t_L^0$ then: defendant chooses $p^c = p*$.
If $t_H^0 > p* \geq t_L^1$ then:

- defendant chooses:
  - $p^c = p*$ if $b(p*) - c(p*) - 5 \geq b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 1$

- defendant chooses:
  - $p^c = t_L^1 - \varepsilon$ if $b(p*) - c(p*) - 5 < b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 1$

If $t_H^1 > p* \geq t_H^0$ then:

- defendant chooses:
  - $p^c = p*$ if $b(p*) - c(p*) - 5 \geq b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 1$
- defendant chooses:
  - $p^c = t_L^1 - \varepsilon$ if $b(p*) - c(p*) - 5 < b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 1$

If $p* \geq t_H^1$ then:

- defendant chooses:
  - $p^c = p*$ if $b(p*) - c(p*) - 10 \geq b(t_H^1 - \varepsilon) - c(t_H^1 - \varepsilon) - 5$
  - and $b(p*) - c(p*) - 10 \geq b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 1$
- defendant chooses:
  - $p^c = t_H^1 - \varepsilon$ if $b(p*) - c(p*) - 10 < b(t_H^1 - \varepsilon) - c(t_H^1 - \varepsilon) - 1$
  - and $b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 5 < b(t_H^1 - \varepsilon) - c(t_H^1 - \varepsilon) - 1$
- defendant chooses:
  - $p^c = t_L^1 - \varepsilon$ if $b(p*) - c(p*) - 10 < b(t_L^1 - \varepsilon) - c(t_L^0 - \varepsilon) - 1$
  - and $b(t_H^1 - \varepsilon) - c(t_H^1 - \varepsilon) - 5 < b(t_L^1 - \varepsilon) - c(t_L^1 - \varepsilon) - 1$

What this means in terms of the pre-2010 crack-cocaine defendants:

1. Some portion of defendants locate in $p\epsilon(0,5)$ range because it is where MC(p) = MB(p). Another portion of defendants locate in that range because the sentencing consequence of being at their preferred **p** is too high. Those defendants should bunch at $p = 5 - \varepsilon$.

2. Similarly, some portion of defendants in $p\epsilon[5,50)$ range are there because it is the preferred **p**, others are there to escape the sentencing consequence of their preferred **p.** Those who choose **p** to avoid the mandatory minimum should be bunched at $p = 50 - \epsilon$

3. Finally, all defendants in the range $p\epsilon[50,\infty)$ are there because it is the preferred **p**.

After 2010, the lower threshold changes from 5g to 28g and the higher threshold changes from 50g to 280g. As a result, defendants should shift:

1. Defendants in $p\epsilon(0,5)$ pre-2010 due to MC(p)=MB(p) will remain in that range. All defendants with preferred **p** between 5g and 28g who bunched below 5g pre-2010 will now locate at their preferred **p**. Some defendants with preferred **p** at or above 28g who bunched below 5g pre-2010 will now bunch below 28g and some will move to their preferred **p**.

2. Defendants in $p\epsilon[5,28)$ pre-2010 are only there because MC(p)=MB(p). No defendants in this range should shift away from the range (but others will move into this range).

3. Some defendants in $p\epsilon[28,50)$ pre-2010 will remain there (if there preferred **p** still dominates bunching below 28g). Some defendants will now bunch below 28g. Finally, some defendants located in $p\epsilon[28,50)$ because it dominated locating at or above 50g and receiving a higher sentence. Some of those defendants are now free to locate at or above 50g and some will bunch below 28g.

4. Defendants in $p\epsilon[50, 280)$ pre-2010 will now bunch below 28g or remain at their preferred **p** in the 50g-280g range.

5. Defendants in $p\epsilon[280, \infty)$ pre-2010 will now bunch below 280g or remain at their preferred **p** at or above 280g.

Empirically, I find no evidence of an offender response to the Fair Sentencing Act in 2010 as predicted above.

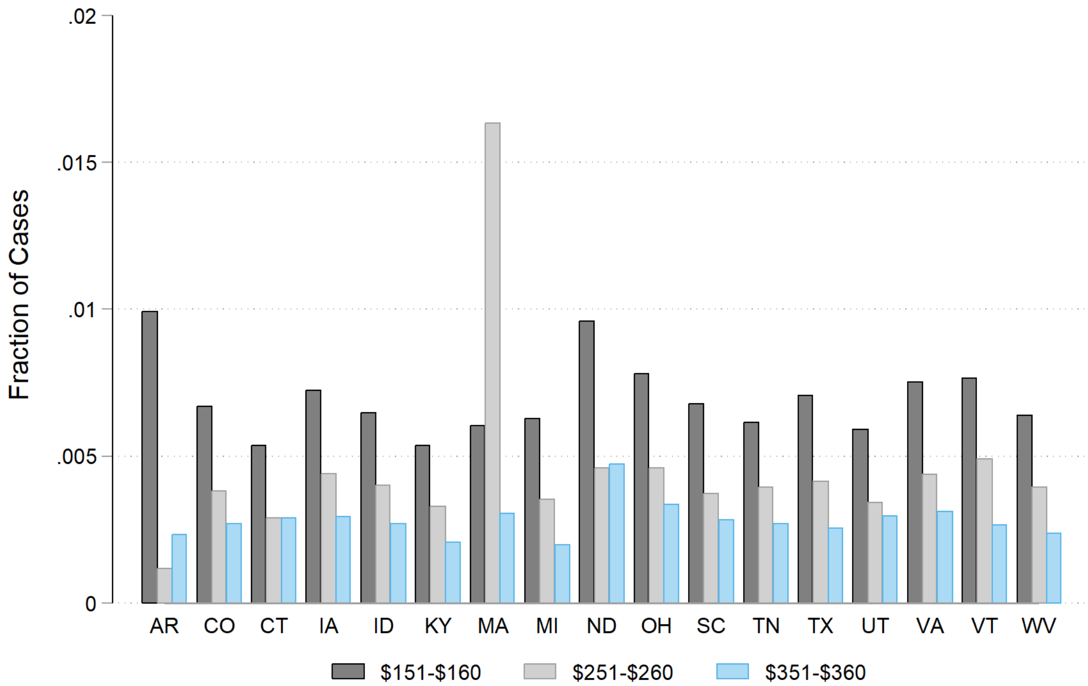
Appendix D. Bunching at Felony Property Crime Thresholds

The value of goods stolen or damaged in a property crime determines whether the crime is a misdemeanor or a felony. Felony property crime carries a harsher penalty, and is typically triggered by a state-level cutoff value. In Massachusetts, property crime that involves goods valued above $250 constitutes a felony. In many other states, this cutoff is around $500 or $1000.

I use data from the FBI's National Incident Based Reporting System (NIBRS) on the amount of goods stolen or damaged in reported property crimes in 2000, 2005, 2010, and 2015. Since property crime thresholds are set at a state-level, I can compare the distribution of the value of goods in Massachusetts (where the threshold is $250) to the distribution in other states (where the threshold is typically $500 or $1000).

Figure D1 shows the fraction of cases with goods valued from $151-$160, $251-$260, and $351-$360 by state in the year 2000. The fraction of cases with goods valued from $251-$260 in Massachusetts is much higher than average, and it is the only state from 2000-2015 that has a felony theft threshold at $251.

Figure D1. Evidence of Bunching at Property Crime Threshold in MA, 2000.



Notes. In the figure above, I plot the fraction of property crime incidents involving goods valued $151-$160 (dark gray), $251-$260 (light gray), and $351-$360 (light blue) for each state in the year 2000. State names are listed on the y-axis. More states begin reporting incident-level information in later years, because of this, the figure above will look similar but more crowded in 2005, 2010, and 2015. In the bunching analysis, I include all states and all years from 2000, 2005, 2010, and 2015.

To calculate the degree of bunching above $250 in Massachusetts, I estimate the following linear probability model:

$$(Recorded\,\$251-\$260)_i = \alpha_0 + \beta_1 MA_i + \epsilon_i \tag{4}$$

where $(Recorded\,\$251-\$260)_i$ is equal to one if the value of goods involved in the offense is recorded as $251-$260 (including $260) and is equal to zero if the value of goods is not in that range. MA_i is equal to one if the offense occurs in Massachusetts and is equal to zero if the offense occurs in any other state in the data. I restrict the sample to crimes with goods valued from $0-$100000 to remove extreme values. I also estimate where the excess mass at $251-$260 comes from with the following model:

$$(Recorded\,\$X-\$Y)_i = \alpha_0 + \beta_1 MA_i + \epsilon_i \tag{5}$$

where the ranges used are $0-$250 (including $250) and $260-$100000 (not including $260). I also show that there is heterogeneity in bunching across agencies in Massachusetts, with some agencies exhibiting no bunching and others exhibiting substantial bunching from $251-$260.

In Table D1, I estimate the degree of bunching above the $250 property crime threshold in Massachusetts by comparing offenses in Massachusetts to offenses in other states. All specifications below use a balanced sample of agencies in each state. Column 1 shows that a property crime offense in Massachusetts is about 1.5 percentage points more likely to be recorded as involving goods valued from $251-$260 than in other states.

Table D1. Effect of Changing Threshold on Bunching at $251-$260.

	Pr(Value of Goods $251-$260) (1)
In Massachusetts	0.0143***
	(0.0001)
Constant	0.0039***
	(0.0001)
Observations	5427406
R-squared	0.0014

Notes. Standard errors clustered at the state-level in parentheses. All specifications above use the sample of offenses with goods valued between $0 and $100000. Data from 2000, 2005, 2010, and 2015 NIBRS Property Segment are used. Only agencies with greater than 500 incidents per year are included. All specifications include year fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Tables D2a and D2b suggest that the excess mass just above $250 in Massachusetts comes from below the threshold rather than above it, again suggesting that law enforcement officers are shading the value of the goods upward. In this case, we can be more certain that law enforcement is the source of the bunching because the data come directly from police departments. In addition, Table D2a shows that the probability the value of good is coded as exactly $250 is higher in Massachusetts than in other states. This suggests that theft/property crime may be a case in which some officers discount the value while others shade up the value of the crime.

Table D2a. Missing Mass in Other Parts of the Distribution of Stolen/Damaged Goods.

	Pr($0-$50)	Pr($50-$100)	Pr($100-$200)	Pr($200-$249)	Pr($250)	Pr($260-$300)	Pr(Above $300)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
In Massachusetts	-0.0604***	-0.0119***	-0.0135***	-0.0051***	0.0108***	0.0010***	0.0647***
	(0.0073)	(0.0017)	(0.0025)	(0.0012)	(0.0002)	(0.0003)	(0.0113)
Constant	0.2256***	0.0899***	0.1419***	0.0650***	0.0182***	0.0120***	0.4436***
	(0.0092)	(0.0023)	(0.0042)	(0.0016)	(0.0005)	(0.0004)	(0.0136)
Observations	5427406	5427406	5427406	5427406	5427406	5427406	5427406
R-squared	0.0015	0.0001	0.0003	0.0001	0.0003	0.0000	0.0017

Notes. Standard errors clustered at the state-level in parentheses. All specifications above use the sample of offenses with goods valued between $0 and $100000.
*** p<0.01, ** p<0.05, * p<0.1

Table D2b. Missing Mass in Other Parts of the Distribution of Stolen/Damaged Goods.

	Pr(Below \$251)	Pr(\$251-\$260)	Pr(Above \$260)
	(1)	(2)	(3)
In Massachusetts	-0.0800***	0.0143***	0.0657***
	(0.0111)	(0.0001)	(0.0112)
Constant	0.5406***	0.0039***	0.4556***
	(0.0134)	(0.0001)	(0.0135)
Observations	5427406	5427406	5427406
R-squared	0.0020	0.0014	0.0017

Notes. Robust standard errors in parentheses. All specifications above use the sample of offenses with goods valued between \$0 and \$100000.
*** p<0.01, ** p<0.05, * p<0.1

To assess the degree of bunching for each police agency in Massachusetts, I define two bunching metrics. The first is a within-agency measure that is motivated by the evidence in Figure 3.

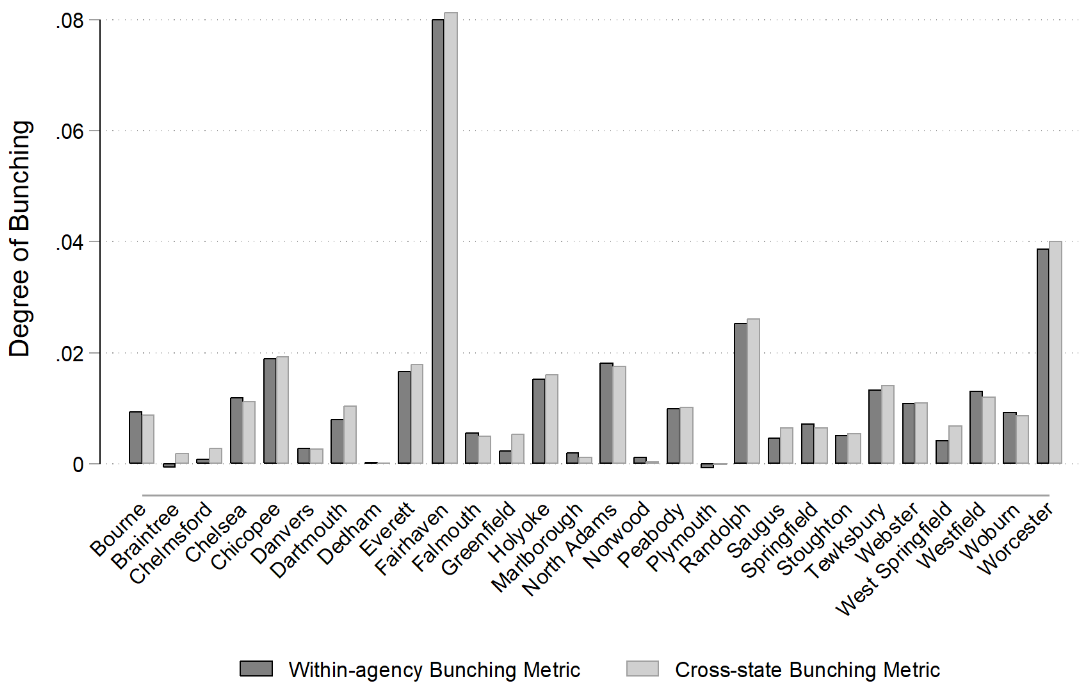
$$\text{Within Agency Bunching}_i = (\text{\% of Cases \$251-\$260})_i - \frac{(\text{\% of Cases \$151-\$160} + \text{\% of Cases \$351-360})_i}{2}$$

where the degree of bunching in agency i is defined as the difference between the fraction of cases in that agency with property values from \$251-\$260 and the average fraction of cases with values from \$151-\$160 and \$351-\$360. The second metric I use is similar to the attorney-level metric I defined in Section III.A.2 above. First, I calculate the mean fraction of cases with values from \$251-\$260 in all states in the data outside of Massachusetts–about 0.004. Then, I subtract that number from the fraction of cases in agency i with values from \$251-\$260.

$$\text{Cross State Bunching}_i = (\text{\% of Cases \$251-\$260})_i\text{-0.04\%}$$

Figure D2 plots both of these bunching statistics for the balanced sample of agencies in Massachusetts. There is substantial heterogeneity in bunching across agencies in Massachusetts. For example, Worcester, Randolph, and Fairhaven exhibit a high degree of bunching above the felony threshold, while Braintree, Plymouth and Dedham exhibit no bunching above the threshold.

Figure D2. Heterogeneity in Bunching by Agencies in MA.



Notes. I plot two agency-level measures of bunching above the felony theft threshold. The data includes all agencies that are present in 2000, 2005, 2010, and 2015 that have at least 500 reported incidents in each year.