Conviction, Incarceration, and Recidivism: Understanding the Revolving Door*

John Eric Humphries, Aurelie Ouss, Kamelia Stavreva, Megan T. Stevenson, Winnie van Dijk†

July 12, 2023

Abstract

We study the effects of conviction and incarceration on recidivism using quasi-random judge assignment. We extend the typical binary-treatment framework to a setting with multiple treatments, and outline a set of assumptions under which standard 2SLS regressions recover causal and margin-specific treatment effects. Under these assumptions, 2SLS regressions applied to data on felony cases in Virginia imply that conviction leads to a large and long-lasting increase in recidivism relative to dismissal, consistent with a criminogenic effect of a criminal record. In contrast, incarceration reduces recidivism, but only in the short run. The assumptions we outline could be considered restrictive in the random judge framework, ruling out some reasonable models of judge decision-making. Indeed, a key assumption is empirically rejected in our data. Nevertheless, after deriving an expression for the resulting asymptotic bias, we argue that the failure of this assumption is unlikely to overturn our qualitative conclusions. Finally, we propose and implement alternative identification strategies. Consistent with our characterization of the bias, these analyses yield estimates qualitatively similar to those based on the 2SLS estimates. Taken together, our results suggest that conviction is an important and potentially overlooked driver of recidivism, while incarceration mainly has shorter-term incapacitation effects.

JEL codes: K42, J24

*Thanks to Meredith Farrar-Owens and others at the Virginia Criminal Sentencing Commission for providing data and answering questions, and to Ben Schoenfeld for web scraping Virginia criminal court records and making them publicly available. We are grateful to Alex Albright, Steve Berry, Jiafeng (Kevin) Chen, Will Dobbie, Deniz Dutz, Brigham Frandsen, Anjelica Hendricks, Felipe Goncalves, Hans Grönqvist, Phil Haile, Randi Hjalmarsson, Rucker Johnson, Larry Katz, Emily Leslie, Charles Loeffler, Jens Ludwig, Alex Mas, Magne Mogstad, Jack Mountjoy, Derek Neal, Arnaud Philippe, Vitor Possebom, Steve Raphael, Yotam Shem-Tov, Elie Tamer, Pietro Tebaldi, Alex Torgovitsky, Crystal Yang, Ed Vytlacil, Chris Walker, and seminar participants for helpful comments. We thank Magdalena Dominguez, Jeff Grogger, Vishal Kamat, and Mike Mueller-Smith for serving as discussants. We thank Naomi Shimberg, Joost Sijthoff, Iliana Cabral, and the UVA Law Librarians for excellent research assistance. We also thank Arnold Ventures and the Tobin Center for Economic Research for financial support. Any remaining errors are our own.

†Humphries: Yale University, Ouss: University of Pennsylvania, Stavreva: Columbia Economics Department, Stevenson: University of Virginia Law School, van Dijk: Harvard University.

1 Introduction

The U.S. criminal justice system is commonly referred to as a "revolving door" due to the high rate of recidivism among those who come into contact with it.[1] A key question for policy-makers is whether the criminal justice system itself contributes to these patterns or whether external factors are responsible, such as addiction, mental health, neighborhood disadvantage, or limited labor market opportunities. Much of the discussion has focused on how *incarceration* affects recidivism. However, for every incarcerated person in 2010, roughly three more were recently *convicted* but *not incarcerated* (Phelps, 2013). A conviction could directly affect recidivism through several channels. It may induce crime by reducing its opportunity cost. For example, a criminal record could make it harder to find employment. Furthermore, a conviction may increase future criminal justice contact even if it has no impact on criminal behavior. For example, prosecutors may be more likely to pursue charges against someone with a record, and judges may sentence them more harshly. Conversely, a conviction could act as a deterrent if it increases the expected penalties for future crime (Drago et al., 2009).

In this paper, we provide new evidence on how both conviction and incarceration affect future criminal justice involvement. Our main approach uses quasi-random assignment of cases to judges as a source of exogenous variation. This approach is similar to existing research studying causal effects of incarceration, but our discussion formalizes an extension of this research design to three, instead of two, possible treatments. Our goal is to estimate *margin-specific* treatment effects, i.e., causal impacts of conviction without incarceration ("conviction") relative to dismissal of all charges ("dismissal"), and causal impacts of incarceration relative to conviction.[2] Such contrasts allow us to isolate the impact of mechanisms that come into play when someone is convicted (such as having a criminal record, or increased supervision), from the impact of mechanisms that matter for incarceration (such as incapacitation).

We study a newly-constructed panel of felony cases in Virginia, spanning approximately two decades. Our outcomes are new felony charges, convictions, and carceral sentences, which, following the literature, we refer to as "recidivism." This could capture both new criminal activity as well as discretionary charging and sentencing decisions. Our results point to conviction as an important, long-lasting driver of future interactions with the criminal justice system, while our analysis of the impact of incarceration only finds evidence for a shorter-term decrease in recidivism (which is likely

[1] According to the Bureau of Justice Statistics, 44% of people released from prison in the U.S. in 2005 were rearrested within one year. Nine years later, 83% had been rearrested at least once (Alper et al., 2018).

[2] We follow the literature in referring to estimands as "causal" if they are a non-negatively weighted average of local average treatment effects (LATEs).

due to incapacitation). Our findings on the impact of incarceration are in line with a substantial portion of the literature, with a recent review concluding that *"Most [quasi-experimental] studies find that the experience of postconviction imprisonment has little impact on the probability of recidivism"* (Loeffler and Nagin, 2022). By contrast, our results for the impact of conviction relative to dismissal point in the opposite direction and are consistent with a criminogenic effect of criminal justice contact, for example due to a criminal record.

Our discussion proceeds in five steps. First, we discuss the interpretation of standard judge-stringency 2SLS estimands in a multiple-treatment setting. We propose a natural extension of the standard IV assumptions in Imbens and Angrist (1994) from the binary case to the case with multiple treatments. Then, we show that, under these assumptions, stringency instruments are not necessarily *treatment-specific*: they can induce cases to switch into more than one treatment.[3] As a consequence, 2SLS with stringency instruments does not generally recover a *causal* effect of a given treatment relative to a weighted average of the other treatments, let alone the *margin-specific* treatment effects.[4] We derive an expression for the (asymptotic) bias that results from this issue, which is additive and easy to interpret. This expression highlights that one of two conditions must be met in order for 2SLS estimands to recover causal and margin-specific effects. One option is to additionally assume homogeneity in treatment effects. The other option is to assume that the instrument moves cases only across a single margin, e.g., from dismissal to conviction. Based on this observation, and drawing on recent literature, we propose a set of assumptions that allow us to interpret the 2SLS estimand as causal and margin-specific.

Second, we estimate the impacts of conviction and incarceration on recidivism using the conventional 2SLS approach. We use the conviction propensity of judges as an instrument for conviction, while controlling for their incarceration propensity to address concerns about violations of the exclusion restriction (since the propensities are correlated).[5] Analogously, we use judges' incarceration propensity as an instrument

[3]Judge stringencies are the *shares* of cases a judge allocates to the different court outcomes. Therefore, being assigned a judge with a higher incarceration stringency will induce some cases to switch into incarceration. But this re-assignment may also cause some cases to switch across the dismissal-conviction margin (even if we condition on conviction stringency). This concern is not present in the judge-stringency framework when treatment is binary. It is also less of a concern in applications where instruments are thought to vary the net payoff to taking up a specific treatment, as in Kirkeboen et al. (2016); Kline and Walters (2016); Mountjoy (2022), since those instruments are likely to only induce switching into only one treatment.

[4]Challenges in obtaining margin-specific treatment effects, even with well-defined causal effects in hand, are discussed in, e.g., Heckman et al. (2006, 2008); Kirkeboen et al. (2016); Kline and Walters (2016); Mountjoy (2022); Heinesen et al. (2022); Bhuller and Sigstad (2022).

[5]This approach mirrors a strategy used in the literature studying the impact of incarceration on recidivism. See Loeffler and Nagin (2022) and Doleac (2023) for recent reviews of this literature. Many papers have used the random assignment of judges to study the impact of court orders and 'examiner' decisions in other settings, including bankruptcy protection (Dobbie and Song, 2015), disability claims (Maestas et al.,

for incarceration and control for their dismissal propensity. Under the assumptions described in the first part of our discussion, our estimates would imply that conviction leads to large and long-lasting increases in recidivism, while incarceration decreases recidivism in the first year, likely due to incapacitation. We estimate that conviction relative to dismissal increases future charges by 11 percentage points in the first year (95% CI, 0.03 to 0.19), 14 percentage points in the first four years (95% CI, 0.00 to 0.28), and 23 percentage points in the first seven years (95% CI, 0.04 to 0.42). In contrast, we estimate that incarceration relative to conviction decreases future charges by 10 percentage points (95% CI, -0.15 to -0.05) in the first year, 8 percentage points in the first four years (95% CI, -0.17 to 0.00), and 7 percentage points in the first seven years (95% CI, -0.19 to 0.05). We find similar results when using future convictions or future incarceration as our measures of recidivism.

Third, we discuss how to weigh this empirical evidence by returning to a key assumption that we invoked to interpret our estimates as causal and margin-specific. We build intuition for its restrictiveness by examining implications for standard index models of judge decision making, and then relating these models to our institutional setting. We show that an ordered model of judge decision making is consistent with the interpretation of the 2SLS estimand as causal and margin-specific, while less restrictive models are generally not. For example, stringency instruments are not treatment-specific under the less restrictive unordered multinomial choice model, where judges consider multiple dimensions of unobserved heterogeneity. Hence, they do not generally allow us to recover causal, let alone margin-specific, estimands, and they are not consistent with the stronger assumptions that we outlined earlier.

Fourth, we propose and implement an empirical test of the assumptions that we invoked to interpret our estimates as causal and margin-specific. Given the preceding discussion, the test also lets us adjudicate between different models of judge decision-making. In particular, we test whether each of the stringency instruments move compliers across a single margin. The test evaluates whether varying an instrument changes the observed characteristics of individuals in the treatment group that is not on either side of the margin of interest. For example, for the dismissal-conviction margin, we evaluate whether the observed characteristics of the incarcerated change when varying conviction stringency and fixing incarceration stringency. We find evidence that, in our context, neither stringency instrument moves people across only a single margin. This implies that we can empirically reject the standard ordered and sequential models of judge decision making, and that our 2SLS estimands are likely (asymptotically) biased. However, returning to the expression we derived for the bias, we use theory and

2013; French and Song, 2014), foster care placement (Doyle, 2008; Bald et al., 2022; Gross and Baron, 2022), evictions (Collinson et al., 2022), or pretrial conditions (Leslie and Pope, 2017; Dobbie et al., 2018).

external evidence to characterize the likely sign and magnitude of the bias, and argue that it is unlikely to be large enough to lead to a qualitatively different conclusion.

Fifth, given that the 2SLS estimates may be biased, we provide two alternative approaches for identifying and estimating the impacts of conviction and incarceration. The first involves a novel approach that builds on Mountjoy (2022) to identify margin-specific treatment effects in a multiple-treatment context. This approach requires treatment-specific instruments, which we have argued judge stringencies generally are not. Following methods from the discrete choice literature, we impose some additional structure on the choice problem to construct treatment-specific instruments from judge stringencies. We then use these treatment-specific instruments to obtain estimates of margin-specific treatment effects. The results are similar to those based on the 2SLS estimates, although they are somewhat smaller and sometimes less precise. The second approach uses a regression discontinuity design which exploits discontinuities in the sentence guidelines to estimate the impact of incarceration relative to conviction. Using two different discontinuities, we evaluate both the extensive and intensive margin of incarceration. Again, results are qualitatively similar to 2SLS: short-term declines in recidivism consistent with incapacitation, and no evidence of longer term effects.

This research contributes to two strands of the literature aiming to understand whether and how the design of the criminal justice system increases or reduces overall crime. First, our work is related to a small set of recent studies that explore the impact of criminal convictions, most of which considers the impact of conviction as part of a broader bundle of treatments. Two recent studies show that felony diversion causes large and sustained reductions in future criminal justice contact (Mueller-Smith and Schnepel, 2021; Augustine et al., 2022). Felony diversion can affect recidivism through several channels in addition to reduced likelihood of conviction, including enhanced deterrence (since rearrest leads to reinstated charges) and lower likelihood of incarceration.[6] Nonetheless, the authors present compelling evidence that felony conviction plays a substantial role in the documented effect. In the context of misdemeanors, Agan et al. (2021a) show that the decision to file charges increases future contact with the criminal justice system. However, only 26% of those charged receive a misdemeanor conviction and the authors argue that the mark of a conviction is not the main channel explaining this effect. Another closely related line of research has documented the consequences of widespread screening of criminal records in labor, housing, and credit markets, suggesting possible mechanisms through which conviction can affect recidivism (e.g., Pager, 2003; Holzer et al., 2006, 2007; Agan and Starr, 2018; Doleac and Hansen, 2020; Craigie, 2020; Rose, 2021a; Cullen et al., 2022; Agan et al., 2022).

[6] Drago et al. (2009) provide evidence that threats of future sentences decrease crimes.

We contribute to this literature both by disentangling conviction from other aspects of the criminal process and by assessing the relative importance of conviction and incarceration in explaining future criminal justice involvement. Our analysis is focused on those facing felony charges, an important category given that 8% of U.S. adults and 33% of Black men have a felony conviction (Shannon et al., 2017). Our findings also relate to past work by sociologists and legal scholars (e.g., Natapoff, 2011; Phelps, 2013, 2017; Brayne, 2014; Kohler-Hausmann, 2018) who suggest that conviction or its components, like a criminal record or supervision, can influence people's future trajectories through changes in behaviors and changes in the outcomes of future criminal justice interactions.

Second, this paper contributes to the large body of work investigating the consequences of incarceration for recidivism and re-incarceration (e.g., Kling, 2006; Hjalmarsson, 2009; Aizer and Doyle, 2015; Kuziemko, 2013; Loeffler, 2013; Mueller-Smith, 2015; Gupta et al., 2016; Leslie and Pope, 2017; Estelle and Phillips, 2018; Harding et al., 2018; Dobbie et al., 2018; Franco et al., 2020; Bhuller et al., 2020; Norris et al., 2021; Rose and Shem-Tov, 2021; Arteaga, 2021; Jordan et al., 2022; Garin et al., 2022). The findings from this literature are mixed, with previous studies finding positive (e.g., Bhuller et al., 2020), negative (e.g., Aizer and Doyle, 2015; Mueller-Smith, 2015), or no evidence of a long-run effect of incarceration on recidivism (e.g., Loeffler, 2013; Rose and Shem-Tov, 2021). We contribute to this strand of literature by highlighting the challenges of studying the impacts of incarceration on recidivism using randomly assigned judges when judges can influence the decision to convict. These challenges may explain some of the mixed results in the literature. In particular, RD evidence tends to find only short-run effects associated with incapacitation, but it exploits discontinuities in sentence recommendations for a sample of people *who are all convicted*. In contrast, the judge IV evidence is mixed, but tends to compare people incarcerated to people facing a mixture of non-carceral outcomes. We characterize when 2SLS will be asymptotically biased, and propose an alternative estimation method. Furthermore, this paper evaluates the impacts of incarceration using two independent sources of variation within the same institutional setting. As in Garin et al. (2022), who similarly use a judge IV and a regression discontinuity research design, our findings about the effects of incarceration across the two research designs are consistent, lending additional credibility to our conclusions.

Our findings are complementary to recent papers that find no long-run effects of incarceration on recidivism. Rose and Shem-Tov (2021) use regression discontinuities that generate variation in exposure to incarceration within a sample of people convicted of felonies. Norris et al. (2021) use a judge IV research design in a setting where the number of people whose cases are dismissed is small. Hence, the counterfactual to incarceration in both papers is typically conviction. We replicate their long-run null findings for recidivism in our setting, but also provide evidence that a felony conviction increases recidivism relative to having the case dismissed. Garin et al. (2022) revisit the research designs in Rose and Shem-Tov (2021) and Norris et al. (2021), and additionally document null effects on long-run labor market outcomes. As highlighted by these authors, their findings suggest that upstream factors such as conviction or socioeconomic disadvantage are likely to explain the tenuous labor market attachments among the formerly incarcerated. Our findings on the role of felony conviction are in line with this conjecture, although our focus is on recidivism rather than labor market outcomes. Our findings also aid in the interpretation of the causal effects in the aforementioned studies, insofar as the null effects could be partially driven by the disruptive effects of a felony conviction that applies to all individuals in their sample.

Lastly, our paper builds on the literature on the identification and estimation of treatment effects in the presence of multiple treatment alternatives and where margin-specific effects are the objects of interest (e.g. Heckman and Vytlacil, 2007; Kline and Walters, 2016; Kirkeboen et al., 2016; Heckman and Pinto, 2018; Mountjoy, 2022; Heinesen et al., 2022; Bhuller and Sigstad, 2022). Lee and Salanié (2018) consider modeling multiple dimensions of unobserved heterogeneity in such settings. The conviction and sentencing decisions we study in this paper are a natural application of this type of framework, since judges may, for example, be influenced by both unobserved strength of evidence and unobserved perceived risk of re-offending.

We also build upon prior work exploring multiple treatments in a judge IV framework, such as Mueller-Smith (2015) and Arteaga (2021). Mueller-Smith (2015) provides one of the first sustained discussions of the challenges to the judge IV research design in a multiple-treatment context and proposes controlling for judge stringency along 'non-focal' dimensions (such as fine amount or probation length). Arteaga (2021) discusses multiple-treatment identification issues within a sequential model. Our work builds on these two papers by laying out identifying assumptions for margin-specific causal effects and providing a more general discussion of which decision-making models satisfy those assumptions. We also focus on conviction as opposed to solely on incarceration. Beyond the criminal justice setting, there are many settings exploited in the literature where examiners choose among more than two options. For example, in foreclosure proceedings, judges can choose between foreclosure sale, loan modification, or case dismissals (Diamond et al., 2020); medical providers can choose to prescribe opioids, other medication, or nothing for pain (Eichmeyer and Zhang, 2022); child welfare investigators choose whether to sustain a claim, and if the claim is sustained, what actions to take (Baron and Gross, 2022). The framework and approach that we develop could be adapted to these cases.

Another closely related paper is Mountjoy (2022), which considers the effects of attending two-year or four-year colleges and proposes an approach to identifying margin-specific treatment effects when there are three possible treatment alternatives. We apply a similar framework in the criminal justice setting. However, an important difference between Mountjoy (2022) and our empirical strategy is that judge stringency instruments are generally not treatment-specific. We propose a way to address this issue by inverting the judge stringency instruments to build treatment-specific instruments. We consider several approaches to implementing the inversions.

The paper proceeds as follows. Section 2 describes the institutional setting and our data. Section 3 extends the random judge design to multiple treatments and discusses when 2SLS recovers causal and margin-specific treatment effects. Section 4 introduces the empirical evidence based on 2SLS estimates. Section 5 discusses the interpretation of the 2SLS estimates under different models of judge decision making, while Section 6 discusses our tests for models of judge decision-making and potential bias. Section 7 describes alternative approaches to identification and estimation, as well as corresponding empirical results. Section 8 provides a discussion that ties together the findings and concludes.

2 Institutional details and data

2.1 Felony case processing in Virginia

This section describes felony criminal case processing in Virginia. It focuses on adjudication within Circuit Court, which is the primary data source for this paper. The flowchart in Figure 1 provides a simplified representation of the process from arrest through disposition.

Between arrest and Circuit Court. After a person is arrested, they are brought to the local police station, booked, and held for their bail hearing. Bail is set by a magistrate: a member of the judiciary who will not preside over further hearings on the case. Charges are first filed in District Court, where the preliminary hearing will be held.[7] At this hearing, the prosecutor must convince the judge that there is probable cause that the defendant committed a felony. This is also the first stage in which plea negotiations might occur. Felony charges might be negotiated down to misdemeanors, or the charges might be dropped or dismissed entirely. If the judge finds probable cause for a felony, the case will then proceed to a grand jury hearing, in which a panel

[7]District Court is a court of limited jurisdiction, meaning that one cannot be convicted of a felony there. District Court adjudicates misdemeanors and provides initial screenings for felonies.

of citizens conduct an additional review of the evidence to ensure that probable cause has been met. If the grand jury finds probable cause that the defendant committed a felony, charges will be filed in Circuit Court, where the remainder of the criminal proceedings will occur.[8] Our analyses focus on cases that make it to Circuit Court (roughly 90% of felony charges).

Assignment of cases to judges. Once charges have been filed in Circuit Court, the case will be assigned to a judge. The exact assignment procedure varies across jurisdictions.[9] A few examples include: (1) the clerk drawing colored stickers out of a can to assign judges; (2) a rotating schedule where a judge will see all cases scheduled for that court during that rotation; (3) assignment of judges to cases based on availability; and (4) cases assigned to judges based on if the case number is odd or even. Appendix E shows that our results are robust to which jurisdictions we include.

Adjudication within Circuit Court. Once a judge has been assigned, the defendant must decide whether she wants to plead guilty or take the case to trial. Since the decision about how to plead depends partly on her expectations of success at trial, we describe the trial process first. Trials in Virginia can be either in front of a judge, which is called a bench trial, or a jury. Approximately 15% of felony convictions in our sample come from trials, almost all of which are bench trials. The remainder come from guilty pleas.[10] In a bench trial, the judge decides whether to convict and, if so, what sentence to give.[11] Judges also exert substantial indirect influence on adjudication and sentencing through various motions. For instance, judges decide what evidence is admissible, what charges can proceed, what must be struck from the record, and what instructions the jury receives. Many of these decisions are made prior to trial. Since they influence the expected outcome of a trial case, they also influence the willingness to offer or accept a plea deal. The more motions are resolved in favor of the defense, the stronger her bargaining position will be. Plea negotiations may result in a stipulated sentence and/or an agreement that the prosecutor will request a particular sentence. Virginia uses a sentence guidelines system, but the judge makes the final decision about the sentence: they have the latitude to reject any negotiated plea deal and to deviate from the sentence guidelines if they provide a written explanation.

These features show that judges influence both conviction and incarceration deci-

[8]There are some potential variations of this process. For instance, defendants can waive their right to a preliminary hearing or a grand jury hearing, and prosecutors can bypass the preliminary hearing and directly indict the case with the grand jury.

[9]We conducted phone interviews with court clerks to determine how cases were assigned to judges.

[10]Plea resolutions are somewhat less frequent in Virginia than in other states. For example, in 2009, nationally, 93% of felony convictions occurred through a guilty plea (Reaves, 2013).

[11]In a jury trial, the jury decides both guilt and sentencing, although the judge can reduce the sentence.

sions in many ways, even if they do not fully control them. This is important for our research design, since we use judge stringencies as instruments in our main analyses.[12]

Virginia's criminal justice system compared to other states. Appendix A compares aggregate statistics of Virginia's criminal justice system to both national averages and statistics for states considered in other recent studies of the impacts of incarceration. Virginia is similar in terms of incarceration and probation rates, and has similar racial and ethnic composition of its incarcerated population. However, it has lower than average parole rates. This is because Virginia adopted "truth in sentencing" for felony convictions starting in 1995, meaning people convicted of felonies have to serve at least 85% of their prison term. As a result, the initial sentence is much more closely linked to time spent incarcerated than in other places.

2.2 How conviction and incarceration may affect recidivism

Conviction. Receiving a conviction instead of a dismissal could increase or decrease recidivism through a number of channels. It could decrease recidivism via deterrence. For example, a person who is convicted but not incarcerated is often placed on probation, which entails additional surveillance and scrutiny, thus increasing the probability of apprehension. It could also raise sentences conditional on conviction, since prior convictions are used to determine recommended sentences. Both of these channels suggest that conviction increases the expected punishment for future offenses, thereby raising the costs of crime (Drago et al., 2009; Philippe, 2020) and dampening recidivism.

Alternatively, felony convictions may increase recidivism due to the stigma and destabilization associated with such records. Employers or landlords conducting background checks may be dissuaded from hiring or renting to someone with a felony conviction, raising the cost of finding work in the formal sector, depressing future wages, and driving those with felony conviction to move into neighborhoods with higher overall crime rates.[13] A prior conviction may also increase our measures of recidivism by changing the outcomes of future criminal justice interactions, even with no changes to future criminal behavior. Our recidivism measures are based on new felony charges,

[12] We provide more institutional details related to the relevance of judge stringency for case outcomes as well as empirical evidence in Appendix D.

[13] Those with a felony conviction are prohibited by law from certain types of employment and from receiving certain public benefits. While arrests that don't lead to conviction can be viewed on background screening reports from private companies, employers are legally prohibited from making decisions on the basis of an arrest record https://www.eeoc.gov/arrestandconviction). Employment background checks submitted to the Virginia criminal records database do not show arrests that didn't lead to a conviction (see VA Code §19.2-389).

convictions and carceral sentences, all of which involve discretionary decisions by various criminal justice actors. A prior conviction may influence these decisions, leading to a "ratcheting up" of penal responses, where each subsequent interaction with the criminal justice system results in more severe consequences. In Section 4.2, we explore the latter two mechanisms by considering recidivism in income-generating and non-income generating offenses as separate categories.

Incarceration. Incarceration could affect recidivism through a variety of channels. It could reduce future criminal justice contact through incapacitation (Avi-Itzhak and Shinnar, 1973).[14] Incarceration could also decrease recidivism through specific deterrence (Zimring et al., 1973; Drago et al., 2009; Jordan et al., 2022). Under this theory, the negative experience of incarceration discourages future criminal behavior. Alternatively, incarceration could increase recidivism because the trauma, disruption, and loss of human capital involved with time behind bars erode a person's capacity to make a living on the legal labor market (Sykes, 1958; Blevins et al., 2010). Crime becomes more attractive as the outside option becomes less lucrative or less accessible. Prison might also expand the criminal network, thus making illicit activity more profitable (Hagan, 1993; Bayer et al., 2009; Stevenson, 2017).

2.3 Data sources, sample construction, and summary statistics

This subsection provides a brief overview of our data as well as sample and variable construction. A much more detailed description can be found in Appendices B.1-B.4. This subsection also presents summary statistics.

Data. Our primary data source for the judge IV analysis in Section 4 comes from Virginia's Circuit Courts. The data was scraped from a publicly accessible website. The Circuit Court data are available from 2000-2020 and cover all of Virginia except Alexandria and Fairfax counties. This data contains information on charges (type and date), on the defendant (gender, race, and FIPS code of residence), and on court proceedings for these cases (type, outcome, and judge). We also use it to construct defendants' recidivism outcomes. We then supplement this data with information on prior felony convictions from the Virginia Criminal Sentencing Commission (VCSC), which covers everyone convicted of a felony in Virginia during the period 1996-2020.

[14] This doesn't mean that incarceration prevents crime, since crime is rampant in jails and prisons (Wolff et al., 2007). However, most within-prison crime is either not reported or is punished using an internal disciplinary system. Generally, only very serious crimes result in new charges.

VCSC data on sentencing is also the primary source for our regression discontinuity design in Section 7.2.

Sample and variable construction. We drop courts where cases are assigned to judges based on judge specialization or some other non-random schema. We also drop courts where there is substantial missing data as well as those with only one judge. Observations are at the case level. We say that a person is "incarcerated" if at least one charge resulted in a carceral sentence. We define a person to be "convicted" if at least one charge led to a sentence, but none resulted in a carceral sentence. Lastly, we say that a person was "dismissed" if all of their charges led to a dismissal (either by prosecution or judge) or an acquittal. Our main measure of recidivism is whether a person has a new felony charge in Circuit Court for an offense that allegedly occurred after the focal disposition date. Our recidivism measure does not include probation revocations, unless these are also accompanied by a new felony charge for a new crime. We calculate recidivism in the first year, the first four years, and the first seven years after a person's initial conviction. In addition, we calculate rates of new convictions and rates of new carceral sentences (both deriving from a new felony charge) within the same time periods.

Summary statistics. Table 1 provides summary statistics for those dismissed, convicted (without incarceration), or incarcerated, respectively. Slightly more than half of the defendants in our sample received a carceral sentence. Among the non-incarcerated cases, about 65% are convicted. The dismissed, convicted, and incarcerated groups are similar in terms of ZIP code-level poverty, but differ demographically. Cases ending in a conviction (without incarceration) are more likely to have female and non-Black defendants. Cases ending in incarceration are more likely to have defendants with prior felony convictions (23%) compared to the convicted and dismissed samples (11% and 14%, respectively). Drug charges are the most common charges for all groups, followed by larceny, assault, and fraud.[15] Appendix Figure E.1 presents disposition types for four common offenses: drugs, larceny, assault, and fraud.[16] While there is variation in the breakdown, all three disposition types exist within offense type, suggesting that there is policy relevance in considering these various case outcomes across many different case types.

[15] This includes offenses like forgery, credit card fraud, or issuance of false checks.

[16] Jointly, they represent 82% of all cases.

3 Extending the random judge research design to multiple treatments

In this section, we discuss an extension of the frequently-used "random judge" framework from the binary case to the case with more than two possible court outcomes. We first propose natural extensions of IV assumptions that are typical in the binary-treatment case, including a "no defiers" assumption that we call Conditional Pairwise Monotonicity (CPM). We then examine what these assumptions imply for the interpretation of estimands obtained from commonly-used 2SLS regressions that instrument for court outcomes using judge stringencies. We argue that, for the 2SLS estimands to be given a causal (let alone margin-specific) interpretation, these assumptions need to be supplemented with either a restriction on the heterogeneity of treatment effects across individuals, or a restriction on how judges make decisions. We then show that the Unordered Partial Monotonicity (or UPM) assumption proposed by Mountjoy (2022) (and nested by CPM) supports an interpretation of the 2SLS estimands as both causal and margin-specific in our setting.

3.1 Notation and regression specifications

We consider a setting where cases can end in one of three mutually exclusive and collectively exhaustive alternatives, denoted by $T \in \{d, c, i\}$, i.e. cases can end in a dismissal (d), conviction without incarceration (c), or incarceration (i). To simplify the discussion below, we further define $T_k = \mathbb{1}\{T = k\}$ as an indicator for the outcome of the case being $k \in \{d, c, i\}$ and $T_{\setminus d} = \mathbb{1}\{T \in \{c, i\}\}$ as an indicator that is equal to one if an individual is either convicted or incarcerated (i.e., not dismissed). Finally, let $\mathbf{T}$ be the vector $(T_d, T_c, T_i)'$.

Both T_c and T_i are likely to be influenced by unobserved factors that affect recidivism, such as the strength of the evidence or the details of the offense or criminal record. Hence, in a regression of any measure of recidivism on these court outcomes, we would be concerned about potential selection bias. To account for this, a common approach is to use judge propensities for specific case outcomes as instruments. Let J denote the identity of the judge randomly assigned to a case. Define incarceration stringency $Z_i = E[T_i|J]$ and let $z_i^j = E[T_i|J = j]$, where $j \in \{1, ..., \mathcal{J}\}$ indexes the judges. Similarly define Z_k and z_k^j for $k \in \{c, d\}$, and let $\mathbf{Z}$ be the vector $(Z_d, Z_c, Z_i)'$.

Using the notation above and abstracting from covariates, the applied literature commonly uses the following specification to study the impacts of incarceration:

$$T_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_d + U \tag{1}$$

$$Y = \beta_0 + \beta_1 T_i + \beta_2 Z_d + V. \tag{2}$$

This 2SLS regression instruments incarceration with the assigned judge's incarceration stringency, and controls for dismissal stringency Z_d to prevent exclusion violations stemming from the judge affecting other aspects of the case (see, for example, Mueller-Smith, 2015; Arteaga, 2021; Norris et al., 2021).[17]

This distinction is important, because we are interested in isolating the impact of the mechanisms that come into play when someone is incarcerated (such as incapacitation or socialization into criminal behavior), separate from mechanisms that play a role in conviction (such as having a criminal record).

In order to learn about the impacts of getting a criminal conviction, one approach is to run a regression analogous to the one above, but where the focal treatment is now conviction and the non-focal treatment is incarceration:

$$T_c = \gamma_0 + \gamma_1 Z_c + \gamma_2 Z_i + U \tag{3}$$

$$Y = \delta_0 + \delta_1 T_c + \delta_2 Z_i + V. \tag{4}$$

As before, we are interested in conditions under which δ_1 has a causal and margin-specific interpretation as the impact of conviction relative to dismissal.

To build intuition, in the next subsections we consider what can be learned from the Wald estimands resulting from simple pairwise judge comparisons when cases are randomly assigned, and when one judge is more stringent in one dimension (while conditioning on the other stringency dimension). This analysis is instructive because, under standard IV assumptions, the 2SLS regressions (1)-(2) and (3)-(4) recover nonnegatively-weighted averages of such Wald estimands, provided the conditioning stringency is included as a sufficiently flexible control (Blandhol et al., 2022).[18]

Before proceeding with our discussion of pairwise judge comparisons in Section 3.3, we discuss our extension of the typical assumptions from the binary "random judge" research design to a setting with multiple treatments.

[17]Another specification encountered in the literature is a 2SLS specification where the second stage includes the two endogenous treatments, which are instrumented with both stringencies. This specification produces the same estimand as (1)-(2) (see Appendix F.1). Alternatively, researchers may instrument a binary treatment indicator (e.g., for incarceration) with judge stringency in that same dimension, omitting controls for other dimensions of sentencing. In our multiple-treatment framework, this does not recover a well-defined causal effect of incarceration relative to a mix of counterfactuals (see Appendix F.2).

[18]As is common in other studies that use random assignment of court cases to judges, we will include place-and-time fixed effects in our regression specifications. We provide the specification we use for estimation in Section 4, but leave this conditioning implicit until then.

3.2 Extending IV assumptions from the binary case to the multiple-treatment case

We define, for each individual, the potential case outcomes $T(z_c, z_i) \in \{d, c, i\}$, and the potential recidivism outcomes $Y(t), t \in \{d, c, i\}$. The standard IV assumptions of exclusion, random assignment, and relevance are easily extended from the binary to the multiple-treatment case.

A1. Exclusion: $Y(t, z_i, z_c) = Y(t) \; \forall \; t, z_i, z_c$.

A2. Random assignment: $Y(t), T(z_c, z_i) \perp\!\!\!\perp Z_i, Z_c \; \forall \; t, z_i, z_c$.

A3. Relevance: $E[\mathbf{ZT'}]$ has full rank.

It is arguably less straightforward to extend the standard binary monotonicity assumption to the multiple-treatment setting. We propose a natural extension below.

A4. Conditional pairwise monotonicity (CPM):

For case outcomes k, l, and m, for all z_k, z'_k, z_l with $z'_k > z_k$ and holding z_l fixed:

i $T_k(z'_k, z_l) - T_l(z'_k, z_l) \geq T_k(z_k, z_l) - T_l(z_k, z_l)$,

ii $T_k(z'_k, z_l) - T_m(z'_k, z_l) \geq T_k(z_k, z_l) - T_m(z_k, z_l)$,

iii $T_l(z'_k, z_l) - T_m(z'_k, z_l) \geq T_l(z_k, z_l) - T_m(z_k, z_l)$, or $T_l(z'_k, z_l) - T_m(z'_k, z_l) \leq T_l(z_k, z_l) - T_m(z_k, z_l)$.

CPM is similar to a "no defiers" assumption in the standard binary setting (Imbens and Angrist, 1994). It imposes two restrictions. First, it guarantees that individuals only (weakly) move in one direction across any margin in response to changes in one instrument, conditional on a value of the other instrument. Second, it guarantees that individuals only move into (and not out of) treatment $T = k$ when the associated instrument z_k increases, thus fixing the direction of movement between T_k and the other two margins.

CPM does not rule out flows across margins that are not adjacent to the treatment corresponding to z_k. For example, consider an increase in conviction stringency z_c while holding incarceration stringency z_i fixed. Under CPM, this (weakly) induces flows from dismissal to conviction and from incarceration to conviction. Since incarceration stringency is fixed, any flow from incarceration to conviction must be matched by a flow from dismissal to incarceration. All three flows are permitted by CPM. However, since CPM prevents "defiers", these three are the only possible cross-margin flows.

At this point, it is worth noting a key distinguishing property of judge stringency instruments. Since T_d, T_c, and T_i represent indicators for a set of mutually exclusive and exhaustive events, the instruments are judge-specific probabilities and add up to one: $z_d^j + z_c^j + z_i^j = 1$. As such, judge stringency instruments vary net probabilities of treatment takeup, as opposed to treatment-specific payoffs. Our attempt to hold fixed other relevant aspects of judge decision-making by conditioning on z_l thus amounts to holding constant the net probability that $T = l$. This in turn implies that, if increasing z_k results in a shift from $T = l$ to $T = k$, there must also be a compensating same-sized shift from $T = m$ to $T = l$ in order to keep the net probability of $T = l$ constant. In other words, even under CPM, *judge stringency instruments are not treatment-specific* as they can induce switching into more than one treatment.[19]

In many other settings, instruments vary a decision maker's cost or payoff to one of their choices. For example, consider the application in Mountjoy (2022) where individuals can choose to go to two-year college, four-year college, or not enroll, and the instruments are distances to the nearest two-year and four year colleges. Varying the distance to a two-year college, conditioning on the distance to four-year colleges, varies the cost of attending a two-year college. Thus, it could induce individuals to switch from not enrolling to a two-year college. It could also induce people to switch from a four-year college to a two-year college. However, unlike when instruments are stringencies, this shift does not have to be matched by an equal number of people shifting from not enrolling to enrolling in a four-year college.

In the remainder of this section, we first consider the implications of A1-A4 for identification of causal effects based on simple pairwise judge comparisons. This discussion will show that, to learn about causal effects, we require either a restriction on treatment effect heterogeneity, or a stronger assumption than CPM on judge decision making.[20] We then introduce one final assumption that allows us to move from the simple pairwise judge comparisons to the 2SLS regression.

3.3 Pairwise judge comparisons under A1-A4

In this section we examine what we can learn from simple pairwise judge comparisons under assumptions A1-A4. We consider what the Wald estimand recovers when Z_c

[19] As we will argue in the remainder of the paper, ruling out compensating flows amounts to an arguably strong restriction on judge decision making.

[20] In Section 5, we discuss which types of models of judge decision making are consistent with CPM, as well as those consistent with more restrictive assumptions, to help gain intuition about the relative strengths of these different restrictions.

can take only two values and Z_i is fixed. In Appendix C, we show how these Wald estimands are aggregated up by regressions such as those in Section 3.1.

To begin, consider increasing conviction stringency from z_c to z_c' while holding incarceration stringency fixed at z_i.[21] Let $\omega_{i\to c}$ represent the proportion of people switching from $T = i$ to $T = c$ in response to the instrument shift. Similarly, allow $\omega_{d\to c}$ and $\omega_{c\to i}$ to represent the proportions of people responding by switching across the other margins. Next, let $\Delta_{i\to c}^{Y_c-Y_i}$ represent the local average $Y_c - Y_i$ treatment effect for those who switch from $T = i$ to $T = c$ when the instrument shifts from z_c to z_c', holding z_i fixed. More generally, $\Delta_{j\to k}^{Y_m-Y_n}$ is the treatment effect of moving from $T = n$ to $T = m$ for the group of people induced to move from j to k when the instrument shifts from z_c to z_c', holding z_i fixed.[22]

Proposition 1: Under assumptions A1-A4 above, the Wald estimand of increasing conviction stringency Z_c from z_c to z_c', while holding incarceration stringency fixed at $Z_i = z_i$, is given by:

$$\frac{E[Y(z_c', z_i) - Y(z_c, z_i)]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} =$$

$$\underbrace{\frac{\omega_{d\to c}\Delta_{d\to c}^{Y_c-Y_d} + \omega_{i\to c}\Delta_{d\to i}^{Y_c-Y_d}}{\omega_{d\to c} + \omega_{i\to c}}}_{\text{Weighted avg. of } Y_c - Y_i \text{ treatment effects}} + \underbrace{\frac{\omega_{i\to c}}{\omega_{d\to c} + \omega_{i\to c}}\left[\Delta_{d\to i}^{Y_i-Y_c} - \Delta_{i\to c}^{Y_i-Y_c}\right]}_{\text{Bias term}}. \tag{5}$$

Proof: See Appendix C.1.

Proposition 1 shows that the Wald estimand can be decomposed into two terms. The first term is a weighted average of two LATEs for conviction versus dismissal for two different groups of compliers. The second term represents the bias. This expression illuminates that when treatment effects are homogeneous (in which case $\Delta_{d\to i}^{Y_i-Y_c} = \Delta_{i\to c}^{Y_i-Y_c}$), or when no compliers switch from incarceration to conviction (in which case $\omega_{i\to c} = 0$), the bias term is zero and we recover the margin-specific causal effects of conviction versus dismissal. When neither of these conditions are met, the Wald estimand does not recover a nonnegatively-weighted average of local average treatment effects. In the next subsection, we consider a condition, proposed elsewhere in the literature, that would set $\omega_{i\to c} = 0$.[23]

[21] We derive the bias for impacts of conviction vs dismissal here; the bias for other margins is analogous and only requires rearranging subscripts.

[22] To keep notation simple, we suppress notation indicating the values of the instruments, for example, we write $\omega_{d\to c}$ rather than $\omega_{d\to c}(z_c', z_c|z_i)$ and $\Delta_{j\to k}^{Y_m-Y_n}$ rather than $\Delta(z_c', z_c|z_i)_{j\to k}^{Y_m-Y_n}$

[23] In Appendix C.2 we derive the bias in a setting with four treatments and four stringency instruments.

In Section 6.1 we return to the expression in Proposition 1, and use it to inform a discussion of what the sign and magnitude of the bias might be in our application with criminal court data.

3.4 Pairwise judge comparisons under stronger assumptions

If we are unwilling to assume treatment effect homogeneity, an assumption stronger than A4 can be invoked that would shut down complier flows across the undesirable margin. This assumption is referred to as unordered partial monotonicity in Mountjoy (2022).[24]

A5. Unordered Partial Monotonicity (UPM):

For all z_k, z'_k, z_l with $z'_k > z_k$ and holding z_l fixed:

i $T_k(z'_k, z_l) \geq T_k(z_k, z_l)$

ii $T_l(z'_k, z_l) \leq T_l(z_k, z_l)$

iii $T_m(z'_k, z_l) \leq T_m(z_k, z_l)$

It is possible that UPM holds when varying one instrument and holding the other fixed, while it does not hold when switching the roles of the two instruments. We therefore use the notation UPM($Z_k|Z_l$). UPM($Z_k|Z_l$) implies that increasing Z_k while holding Z_l fixed weakly induces some individuals into $T = k$ from at least one of the other two treatments, and does not induce anyone out of $T = k$. Moreover, the second and third inequalities imply that increasing z_k cannot cause individuals to switch into $T = l$ or $T = m$. Thus, instruments that satisfy UPM are *treatment-specific*: they can induce complier flows into only one treatment.[25]

Proposition 2: Under assumptions A1-A3, and A5, the Wald estimand of moving judge Z_c from z_c to z'_c while holding $Z_i = z_i$ is given by:

$$\frac{E[Y(z'_c, z_i) - Y(z_c, z_i)]}{E[T_c(z'_c, z_i) - T_c(z_c, z_i)]} = E\left[Y(c) - Y(i) | T_c(z'_c, z_i) = 1, T_d(z_c, z_i) = 1\right] \quad = \Delta^{Y_c - Y_d}_{d \to c}. \tag{6}$$

Proof: See Appendix C.1.

We show that, like above, the Wald estimand is a positively weighted average of margin-specific LATEs for the same margin plus additive bias terms.

[24] Heckman and Pinto (2018) provide a similar "unordered monotonicity" condition in a more general setting with more than three treatments.

[25] Note that the UPM($Z_k|Z_l$) assumption also guarantees that CPM holds, as the only flows of individuals are from $T = l$ to $T = k$ or $T = m$ to $T = k$.

With the stronger UPM assumption, the Wald estimand recovers the margin-specific treatment effect of conviction vs dismissal. This result stems from the UPM assumption combined with the use of judge stringencies as instruments. UPM($Z_c|Z_i$) implies there can be no flows between $T = d$ and $T = i$, while fixing the incarceration stringency implies that the share of people incarcerated cannot change.[26] In combination, these imply that there can be no flows from $T = i$ to $T = c$, so there can only be flows from dismissal ($T = d$) to conviction ($T = c$). This implies that $\omega_{i \to c} = 0$, which simplifies Equation 5 to Equation 6 above. Thus, under the UPM assumption, the Wald estimand recovers a margin-specific LATE. In Appendix C.4, we show how UPM can be relaxed by extending the idea of average monotonicity (Frandsen et al., 2023) to the multiple treatment context.

The Wald estimands above consider only two values of Z_c and condition on Z_i. Following Blandhol et al. (2022), 2SLS specifications that control for (rather than condition on) Z_i recover positively weighted averages of these Wald estimands as long as an additional "rich covariates" assumption holds.

A6. Rich covariates: The linear projection of Z_c on Z_i is equal to $E[Z_c|Z_i]$.

Under A1-A3 and A5-A6, the 2SLS specification shown in Equations (3)-(4) yields a positively weighted sum of the Wald estimates shown in Proposition 2.[27]Assumption A6 can be relaxed by controlling for Z_i more flexibly, or eliminated entirely if the data accommodates a saturated set of fixed effects for Z_i. A more detailed discussion and robustness checks can be found in Appendix C.5.[28]

In summary, under UPM or constant effects, the 2SLS estimand can be interpreted as a positively weighted average of margin-specific effects for compliers as long as A1-A3 and A6 also hold.[29] As we discuss in Section 5, UPM can be a strong assumption in our setting, and only holds in some (arguably restrictive) models of judge decision-making. In Section 6 we discuss testable implications of this assumption and characterize the bias when it doesn't hold. In Section 7, we propose ways of proceeding when the assumption is not met.

[26]Note that it is the combination with judge stringency instruments that makes UPM sufficient to recover margin-specific causal effects. For example, in Mountjoy (2022), UPM is only sufficient for 2SLS to recover the effect of one treatment vs a mixture of next-best options.

[27]Note that assumptions A1-A3, and A5 imply the other assumptions needed in Blandhol et al. (2022) for 2SLS to produce causal estimands. In particular, A5 implies their "Ordered strong monotonicity" (OSM).

[28]Also see Appendix C.5 for a discussion of when more covariates are included in the regression specifications, as they commonly are in practice and as they are in our empirical implementation described in Section 4.

[29]This is also true under CPM (A4) when certain additional restrictions are met. For example, under A1-A4 and A6, 2SLS recovers the causal effect of conviction vs dismissal if the effect of incarceration vs conviction is constant.

4 Conviction, incarceration, and recidivism: 2SLS estimates

4.1 Regression specifications for estimation

Using leave-one-out estimates of judge stringency as our instruments, we consider the following 2SLS regression that is common in the literature (reported here for conviction, the incarceration one being analogous):

$$T_c = \delta_0 + \delta_1 Z_c + \delta_2 Z_i + {\delta_3}' X + U \tag{7}$$

$$Y = \gamma_0 + \gamma_1 T_c + \gamma_2 Z_i + {\gamma_3}' X + V. \tag{8}$$

Here, Y is one of the measures of recidivism described in Section 2.3. The vector X includes court-by-year, court-by-month-of-year, and day-of-the-week fixed effects, as well as controls for offense type, race, gender, and a flag for prior felony convictions. For our measure of judge stringency, we use the tri-yearly leave-one-out conviction and incarceration rates for the judge handling the case.[30] We run these 2SLS regressions on the sample described in Section 2.3.

As discussed in Section 3, under A1-A3, UPM, and sufficiently rich controls, these regression estimates can be interpreted as causal and margin-specific.[31] Specifically, γ_1 can be interpreted as a properly-weighted average of the causal impacts of conviction versus dismissal for compliers.[32]

In Appendix D we discuss how A1-A4 are supported by features of the institutional environment, and provide empirical evidence, based on a standard battery of tests, to help assess their validity. For both the conviction and incarceration regressions, we have a strong first stage with F-statistics of 198 and 386 respectively (Table 2). Figure 2 plots the variation in residualized judge conviction and incarceration stringency, showing that there is substantial variation in each. Appendix Figure D.1 additionally shows that there is also substantial variation in Z_c conditional on Z_i, and vice versa. For balance, Table 3 shows that, while case characteristics are strong predictors of conviction and incarceration, they largely do not predict judge stringencies. For the few covariates with statistically significant loadings, the predicted difference in stringency

[30] We choose a tri-yearly specification to allow for a large number of cases per judge, without requiring that judges behave identically for their entire tenure. We exclude cases assigned to judges who see fewer than 100 cases in the 3-year period.

[31] See Appendix C for additional discussion of what 2SLS identifies when including controls based on Blandhol et al. (2022), and details on the assumption of sufficiently rich controls.

[32] Under CPM rather than UPM, γ_1 is a properly weighted average of the biased Wald estimands derived in the previous section, and therefore does not recover a margin-specific effect without additional assumptions (such as constant treatment effects for the incarceration versus conviction).

tends to be very small (0.02 to 0.04 standard deviations of the residualized stringency measure). To test the exclusion restriction, we show that estimates remain largely unchanged when including sentence-length stringencies as additional controls. Finally, we provide a test of the "no defiers" assumption that is part of both CPM and UPM, with Tables D.2 and D.3 reporting split-sample monotonicity tests and finding the same sign for the first stage across various splits of the data. We postpone discussion and implementation of an additional test of the UPM assumption to Section 6.

4.2 Conviction

Table 4 presents 2SLS estimates of the model in Equations (7)-(8). If given a causal and margin-specific interpretation, these estimates would represent the impact of conviction (without incarceration) on recidivism relative to dismissal for those near the margin. We consider three measures of future criminal justice contact: new felony charges in Circuit Court, a new conviction resulting from felony Circuit Court charges, or a new carceral sentence resulting from felony Circuit Court charges. We use various time windows to measure recidivism, all measured from the time of disposition: year 1, years 2-4, years 5-7, cumulatively for the first 4 years, and cumulatively for the first 7 years. For each of these outcomes, we present OLS and 2SLS regressions.[33]

As discussed in Section 2.2, a conviction (instead of a dismissal) could increase or decrease recidivism through a number of channels, and the sign of the net effect is not clear a priori. If given a causal and margin-specific interpretation, our 2SLS estimates suggest that conviction increases future criminal justice contact relative to dismissal. The estimates for future charges within the first year after conviction are large: around 11 percentage points (95% CI, 0.03 to 0.19). The impacts on cumulative recidivism 1-4 and 1-7 years later are also statistically significant and grow over time, with estimates of 14 percentage points (95% CI, 0.00 to 0.28) and 23 percentage points (95% CI, 0.04 to 0.42) respectively. Results are similar for the other measures of recidivism we consider (future convictions and future incarceration) with similarly-sized point estimates and confidence intervals. We note that despite being statistically distinguishable from zero, the confidence intervals leave room for a fairly wide range of values, as is typical for judge IV research designs. Most of the cumulative effects appear to be driven by the first year, with the estimates for years 2-4 and 5-7 also being positive, but smaller and not statistically significant.

Our 2SLS estimates are similarly signed but substantially larger in magnitude than the OLS estimates. The OLS estimates may suffer from omitted variable bias. One im-

[33]Appendix Table E.1 presents reduced-form estimates. The OLS estimate is from a regression of recidivism on a conviction indicator that is one if the individual is convicted or convicted and incarcerated, and controls for an incarceration indicator.

portant omitted variable is the strength of the evidence, which often consists primarily of witness testimony. Graef et al. (Forthcoming) show that witness appearance in court is by far the most predictive factor in whether the defendant will be convicted. Thus, the sign of the bias in the OLS estimates depends in part on the relationship between witness appearance and the defendant's risk of recidivism. These could be positively correlated if, e.g., witnesses are more invested in securing punishment for high recidivism defendants. Or they could be negatively correlated if, e.g., witnesses are scared of testifying against high recidivism defendants. The fact that victims and bystander witnesses often come from the same socioeconomic groups as defendants also suggests a negative correlation. The same factors that give someone a high recidivism potential – poverty, social marginalization, etc. – may also make it harder for the witnesses to take time off work for a court date, or make them less willing to cooperate with a system they distrust. If so, recidivism will be negatively correlated with conviction and OLS estimates will be downward biased.

Alternatively, IV compliers may be more impacted by conviction than the average defendant. In Appendix Tables E.8-E.9, we show that compliers have a similar racial composition as the overall sample, but are less likely to be in court for violent offenses or "other offenses," and less likely to have a prior conviction.[34]

We provide some analysis to explore whether our results are coming from destabilization or ratcheting up of future criminal justice interactions – mechanisms we discussed in Section 2.2. If conviction makes it harder to find employment due to the mark of a felony record, we might expect to see a more pronounced increase in income-generating crime. We test for this in Appendix Table E.3 and find similar point estimates across income generating and non-income generating crime, but the confidence intervals are too large to draw a firm conclusion. Alternatively, if the ratcheting up effect is operative, conviction may have a larger effect on the more downstream measures of future criminal justice contact, such as future conviction or incarceration, since there have been more discretionary choices along the way. Comparing the three measures of recidivism in Table 4, the point estimates are larger relative to their untreated means for outcomes with more discretionary decisions. These analyses provide no conclusive evidence for or against destabilization or ratcheting up. Yet, while these two mechanisms point to somewhat different interpretations, both imply that conviction can trap a person in the revolving door of criminal justice, increasing not just future charges and convictions, but also future incarceration.

[34]Here, "other offenses" are defined as having any charges related to kidnapping, miscellaneous, sex offenses, or traffic.

4.3 Incarceration

Table 5 presents 2SLS estimates of the model in Equations (7)-(8), but instrumenting for incarceration with incarceration stringency and controlling for dismissal stringency. If given a causal and margin-specific interpretation, these estimates represent impacts of recidivism relative to conviction without incarceration for those near the margin.

We find that incarceration causes a decline in recidivism in the first year after sentencing. Our 2SLS estimates suggest a 10 percentage point reduction in future charges (95% CI, -0.15 to -0.05). The negative estimate in the first year is likely due, at least partially, to incapacitation. While people are incarcerated, new crimes are usually addressed with internal sanctions and are unlikely to result in new felony charges. However, in the medium to long run, we find no evidence that incarceration affects future criminal justice interactions beyond the impact in the first year, with estimates shrinking to an eight percentage point reduction in the first four years (95% CI, -0.17 to 0.00) and a seven percentage point reduction in the first seven years (95% CI, -0.19 to 0.05). Results are similar for future convictions or future incarceration. The point estimates for incarceration are negative, but not always statistically significant. Even when not statistically significant, we can rule out moderate increases in recidivism in the long run. For example, for new charges, we can reject increases in recidivism larger than 2.7 percentage points with 95% confidence.

These findings align with the conclusion drawn in a recent literature review that most of the papers that find incarceration to be criminogenic are looking at pretrial detention, rather than post-sentencing incarceration (Loeffler and Nagin, 2022). Since pretrial detention also increases the probability of conviction (Dobbie et al., 2018; Leslie and Pope, 2017), these papers are effectively estimating the joint effect of conviction and incarceration. In contrast, most papers evaluating the impact of post-conviction incarceration do not find evidence of effects lasting beyond the incapacitation period. Incarceration may be a traumatic event, but most studies find no evidence that it is an important contributor to the revolving door.

4.4 Robustness and subgroup analyses

In this section, we provide a brief overview of robustness tests that are detailed in Appendix E. Our results are robust to our choice of sample restrictions and controls, as shown in Appendix Figures E.2-E.7, which display similar point estimates with similar levels of precision across specifications. Additionally, Appendix Figure E.8 demonstrates no differential mobility out of Virginia based on incarceration outcomes.[35]

[35] We are unable to study differential mobility out of Virginia due to conviction, as less information about defendants is collected for cases ending in dismissal, prohibiting linkage to data on out-of-state moves.

Appendix Tables C.3 and C.4 additionally show that the estimates are robust to the choice of fixed effects and to how flexibly we control for the other stringency.

Next, we explore heterogeneity based on observable characteristics (Appendix Tables E.5 - E.7). First, we consider heterogeneity by priors. Avoiding a first felony conviction might play a pivotal role in people's future trajectories. For example, prospective employers may screen more on the presence of a prior conviction, rather than its recency. Therefore, avoiding initial convictions might be especially important. Likewise, a person's initial carceral experience might be especially destabilizing or, on the contrary, it may uniquely act as a deterrent, as found in Jordan et al. (2022). Appendix Table E.5 shows 2SLS estimates for those with and without prior convictions in the last five years. We find that people without a recent felony conviction have large and sustained increases in recidivism as a result of a felony conviction. Yet, we cannot reject that these estimates are equal to estimates for those with a recent felony conviction, for whom estimates are imprecise—likely because they make up only 20% of the sample. For incarceration, we find that those with and without a prior conviction have similar patterns: short-term incapacitation effects but no detectable long-term effects.

We also explore heterogeneity across race and ZIP code income level. We find qualitatively similar patterns across Black and non-Black defendants. We find suggestive evidence that the estimates are larger for people living in ZIP codes with above-median poverty rates. This could be because felony convictions have more consequences for poorer people, perhaps because a felony conviction shuts out access to housing or other social services.

Overall, these results may be relevant to a number of policy debates. Given a causal interpretation, they suggest that incarceration does not play a prominent role in driving the revolving door, consistent with the recent literature review of Loeffler and Nagin (2022). Those interested in closing the revolving door may want to focus on either reducing the number of people with a felony record, or reducing the impact of such a record. Approaches could involve increasing diversion rates or favoring misdemeanor over felony conviction. Alternatively, they could involve making criminal records easier to seal or regulating their uses in labor market and housing decisions. Of course, all of these actions entail tradeoffs. As just one prominent example, Agan and Starr (2018) find that "ban the box" policies that restrict employers from asking job applicants about criminal histories increased the gap in callback rates between White and Black men.

5 CPM and UPM as restrictions on judges' decision models

In this section, we lay out three natural choices for models of judge decision making, and use these models to discuss what types of behaviors UPM rules out. Throughout this section we use "models of judge decision-making" as a shorthand, although, as we discuss in Section 2, in practice court outcomes reflect a combination of decisions by multiple actors. We discuss three index-crossing models of judge decision making based on three canonical models of multinomial discrete choice: an ordered choice model, a sequential choice model, and an unordered choice model. All three models satisfy the CPM assumption. Only the most restrictive model satisfies the UPM assumption for both instruments, ruling out some arguably realistic judge behaviors. The sequential model illustrates that UPM may be satisfied for one of the instruments but not the other, hence we may be able to recover one margin-specific effect but not the other. For each model we also discuss how it relates to the legal and institutional practices of the criminal proceedings.

5.1 Ordered choice

First, we consider a straightforward extension from the binary index-crossing model, shown to be equivalent to the standard binary LATE model in Vytlacil (2002), to a trinary model: an ordered choice model with a single dimension of case-specific unobserved heterogeneity W. Each judge has their own thresholds for the values of W that would result in dismissal, conviction, and incarceration:

$$\begin{aligned} T_d &= \mathbb{1}\{W < \pi_c(Z_d)\}, \\ T_c &= \mathbb{1}\{\pi_c(Z_d) \leq W < \pi_i(Z_i)\}, \\ T_i &= \mathbb{1}\{W \geq \pi(Z_i)\}, \end{aligned} \tag{9}$$

where $\pi_c(Z_d) \leq \pi_i(Z_i)$ for all Z_d and Z_i. The first panel in Figure 3 visualizes, for two different judges, the regions of W under which each judge dismisses, convicts, and incarcerates. In this example, judge 1 has higher cutoffs for both conviction and incarceration than judge 2.

In an ordered choice model, we can estimate margin-specific treatment effects for both the $T = c$ vs $T = d$ margin and the $T = i$ vs $T = c$ margin. To illustrate this, consider panel (b) of Figure 3, in which both judges have the same incarceration threshold, but judge 2 has a lower conviction threshold, meaning they convict more and dismiss less than judge 1. This figure demonstrates a key point: fixing Z_i and increasing Z_c will result in holding $\pi_i(z_i)$ fixed and decreasing $\pi_c(z_d)$. The only people who will switch treatment status as a result are those who move from $T = d$ to $T = c$. When correctly conditioning, the instruments are treatment-specific, since fixing Z_i and increasing Z_c will induce flows into only one choice ($T = c$) and not into any other treatment. Moreover, the instruments only move individuals across a single margin (from $T = d$ to $T = c$). Similarly, we can learn about the effect of incarceration vs conviction using variation in Z_i and fixing Z_d, for those near the margin. Thus, this choice model satisfies the unordered partial monotonicity assumption for both margins (i.e. UPM($Z_d|Z_i$) and UPM($Z_i|Z_d$) hold).

This model would be appropriate if all judges consider only a single dimension of unobserved heterogeneity in their decision, and they agree on how cases are ranked according to this dimension. The only ways in which judges are allowed to differ in their decision making is by setting different thresholds for assigning cases to each of the outcomes. In practice, however, judges may take into account more than one scalar measure of unobserved heterogeneity. For example, judges may consider both strength of evidence and risk of recidivism when the defendant is not incarcerated. In the remainder of this section, we consider models that allow for multiple dimensions of unobserved differences between defendants.

5.2 Sequential choice

Next we consider a sequential choice model in which the court process consists of two decisions: (1) a dismissal decision and, if not dismissed, (2) an incarceration decision. This reflects the two-step process of criminal cases: a trial to adjudicate guilt or innocence, followed by a sentencing hearing if the person is found guilty. The model allows for judges to consider different, though potentially correlated, unobserved factors in each decision. For example, conviction decisions may depend on the strength of the evidence, which is not observed in our data, while incarceration decisions may depend on other aspects, such as the propensity to re-offend or severity of the crime, which are also not observed in our data.

We can write this as an index model:

$$\begin{aligned}
T_d &= \mathbb{1}\{U_c < \pi_c(Z_d)\} \\
T_c &= \mathbb{1}\{U_c \geq \pi_c(Z_d), U_i < \pi_i(Z_i, Z_d)\} \\
T_i &= \mathbb{1}\{U_c \geq \pi_c(Z_d), U_i \geq \pi_i(Z_i, Z_d)\}.
\end{aligned}$$

In this model, the first choice is between $T \in \backslash d$ (not dismissed) and $T = d$ and depends on unobservable U_c. For cases that switch from dismissed to "not dismissed," there is then a second choice: conviction (without incarceration) or incarceration.

This choice depends on unobservable U_i which can be correlated with U_c.[36] This model is consistent, for example, with only a subset of the information available to the judge being used in each of the two steps (for example, if the first choice depends on the strength of the evidence, while the second depends on risk of crime or perceived appropriate punishment for the crime). It is also consistent with new information arriving at the incarceration stage, such as letters of support for the person convicted of the crime or victim impact statements.

Under the assumptions of the sequential model, it is possible to use 2SLS and the stringency instruments to recover margin-specific treatment effects between $T = i$ and $T = c$, but not between $T = c$ and $T = d$ or $T \neq d$ and $T = d$. Figure 4 illustrates this point. Panel (a) visualizes one judge's decision regions based on U_c and U_i. Panel (b) then compares two judges who have the same probability of dismissal, but where the second judge has a higher probability of incarceration. Here, variation in Z_i holding Z_d fixed induces only changes in court outcomes from $T = c$ to $T = i$ for a set of compliers.

In contrast, panel (c) compares two judges who have the same probability of incarceration (Z_i), but where judge 2 has a lower probability of dismissal (Z_d). Recall that Z_i is the *proportion* of cases that a judge incarcerates. In this figure, Z_i is represented by the fraction of people in the top-right quadrant. For two judges to have the same incarceration stringency, both π_i and π_c must differ across these judges. This comparison then induces three sets of compliers, those moving from $T = d$ to $T = c$, those moving from $T = d$ to $T = i$, and those moving from $T = i$ to $T = c$. This example satisfies CPM since there is only a one-way flow across any given margin. But the flows from $T = d$ to $T = i$ mean that the instrument is not treatment-specific, and UPM($Z_c|Z_i$) is not satisfied.

While the sequential model captures the two-step nature of the criminal proceeding, it may not be a good model if case outcomes are determined by a *joint consideration* of the two dimensions, as may be the case when plea bargaining occurs. In the next subsection, we consider a multinomial choice model, which also has two dimensions of unobserved heterogeneity but allows for both unobservables to affect crossing either margin. This may better capture the intertwined decisions that are common in Virginia and other US jurisdictions due to plea bargaining.

[36] See Heckman et al. (2016) for details on estimating treatment effects in this type of sequential choice model, and Arteaga (2021) for a criminal court application studying the impacts of incarceration using a model similar to the sequential model described above.

5.3 Unordered multinomial choice

In an unordered multinomial choice model, treatment depends on the judge's preference for conviction without incarceration ($\pi_c(Z_c, Z_i)$), the judge's preference for incarceration ($\pi_i(Z_c, Z_i)$), and two scalar unobserved characteristics (V_c and V_i). This model might be an appropriate way to accommodate aspects of the plea bargaining process, since it models case outcomes as being determined by a joint consideration across the two unobserved dimensions. In a plea deal, a defendant typically agrees to plead guilty in exchange for a lower sentence, thus blurring the line between the conviction and sentencing decisions; unobserved determinants of the sentencing decision may be used in the decision to plea guilty.[37] The unordered multinomial choice model can also be written as a latent threshold crossing model:

$$\begin{aligned}
T_d &= \mathbb{1}\{V_c < \pi_c(Z_c, Z_i), V_i < \pi_i(Z_c, Z_i)\} \\
T_c &= \mathbb{1}\{V_c \geq \pi_c(Z_c, Z_i), V_c - V_i \geq \pi_c(Z_c, Z_i) - \pi_i(Z_c, Z_i)\} \\
T_i &= \mathbb{1}\{V_i \geq \pi_i(Z_c, Z_i), V_i - V_c \geq \pi_i(Z_c, Z_i) - \pi_c(Z_c, Z_i)\}.
\end{aligned} \tag{10}$$

In this model, the instruments are not treatment-specific. For example, the propensity of a judge to convict depends on both their preference to incarcerate π_i and their preference to convict π_c, neither of which are directly observed. Panel (a) of Figure 5 visualizes the court outcomes and how they depend on judge preferences and the two unobservables.

For this model, it is not possible to use 2SLS and the stringency instruments to recover margin-specific or treatment-specific treatment effects. To see this, panel (b) of Figure 5 shows how treatments change when holding Z_i fixed and increasing Z_c. In this case, individuals shift from incarcerated to convicted and from dismissed to convicted but, in order to hold the probability of incarceration (Z_i) constant, individuals also shift from dismissed to incarcerated. This flow from dismissal to incarceration violates UPM and demonstrates that instruments neither move individuals into a single treatment nor across a single margin. Results are similar when holding Z_c (or Z_d) fixed and varying Z_i.

These results differ somewhat from Kirkeboen et al. (2016) and Kline and Walters (2016) who show that, under unordered multinomial choice, treatment-specific instruments can provide a weighted average of a given outcome versus the various next-best options. The difference stems from the fact that stringencies as instruments are not generally treatment-specific; judge stringencies for conviction, incarceration, or dismissal do not shift a specific π.[38] If we could shift π_c directly, then decreasing π_c

[37] Alternatively, the defendant negotiates for a lower charge which usually comes with a lower sentence.

[38] Judge stringencies are the *shares* of cases a judge allocates to mutually-exclusive treatments, and shifting

holding π_i constant would result in flows into conviction from the other two treatments and no flows between incarceration and dismissal, as shown in panel (c) of Figure 5. Such a shift would allow us to estimate the impacts of conviction versus a mixture of next-best options as in prior work (e.g., Kline and Walters, 2016; Kirkeboen et al., 2016; Mountjoy, 2022). Given that π_c and π_i are not observed, we instead can only shift or condition on Z_c and Z_i, resulting in variation that violates UPM and does not solely shift people into or out of a particular choice.

In the three choice models above, we made no assumptions on how treatment effects depend on unobserved heterogeneity. If we were willing to assume that treatment effects are homogeneous, then it is possible to use stringency instruments to identify the margin-specific effect between any two of the three margins, regardless of which of the three models generates the data.[39]

6 Testing for and characterizing bias in the 2SLS results

In this section, we first describe and implement an empirical test for whether the UPM assumption holds. By extension, this allows us to adjudicate between the models we consider in Section 5. We then use theory and external evidence to discuss the likely magnitude and direction of the bias in our context.

6.1 Testing the UPM assumption

As discussed above, the UPM assumption could be considered restrictive and rules out certain models of judge decision making. In particular, UPM($Z_c|Z_i$) implies that, when shifting Z_c and holding Z_i fixed, nobody enters or exits incarceration ($T = i$). This has testable implications:

(1) Under UPM($Z_c|Z_i$), the observable characteristics of those with $T = i$ should not change when holding Z_i constant and varying Z_c.

(2) Under UPM($Z_i|Z_d$), the observable characteristics of those with $T = d$ should not change when holding Z_d constant and varying Z_i.

a share does not necessarily correspond to shifting a specific π, even when conditioning on one of the other stringencies. By contrast, in Kirkeboen et al. (2016), Kline and Walters (2016), and Mountjoy (2022) the instruments vary the net payoffs to take-up of one treatment, holding fixed the net payoffs to other choices. Hence, they only induce flows into one specific treatment.

[39] Note that this allows for selection on level (e.g. individuals more likely to recidivate may be more likely to be incarcerated), but not selection on the treatment effect of conviction or incarceration.

If instrumental variation is only causing flows between two treatments, there should be no movement in or out of the third treatment. For example, consider the set of people who are incarcerated in the ordered model when judge incarceration stringency is fixed at Z_i. When holding incarceration stringency fixed in the ordered model, varying conviction stringency Z_c will move people between dismissal and conviction, but will not move people in or out of incarceration. This implies that the observed characteristics of incarcerated individuals should not change, and motivates the first testable implication above. If the characteristics of incarcerated individuals do change, then there must be flows in and out of incarceration, which implies that the instrument is moving people across more than one margin. More generally, this would imply that $\text{UPM}(Z_c|Z_i)$ is violated, as the UPM assumption plus stringency instruments (and the other IV assumptions) insures compliers move across only one margin. Thus, the first testable implication above tests whether $\text{UPM}(Z_c|Z_i)$ holds. By a similar argument, the second testable implication tests whether $\text{UPM}(Z_i|Z_d)$ holds. Since these conditions hold in some of the models we consider, these tests also allow us to reject certain models. In particular, (1) and (2) above must hold for the ordered model, and (2) must hold for the sequential model.

We implement our test using predicted recidivism: an index constructed by regressing recidivism on individual and case characteristics.[40] We test implication (1) by regressing predicted recidivism on our conviction instrument, restricting the sample to those incarcerated and controlling for the incarceration instrument and court-by-time fixed effects. Similarly, we test implication (2) by regressing predicted recidivism on the incarceration instrument, restricting to the dismissed sample and controlling for the dismissal instrument and court-by-time fixed effects. Results are shown in Appendix Table E.10, where Panel A presents tests for (1) and Panel B tests for (2).

Using the predicted recidivism index, we reject $\text{UPM}(Z_c|Z_i)$ and $\text{UPM}(Z_i|Z_d)$, which also means we reject both the ordered and sequential models. For (1) we find that predicted recidivism for the incarcerated group increases with the judge's conviction propensity, holding incarceration propensity constant. For (2) we find that the predicted recidivism for the dismissed group decreases with the judge's incarceration propensity, holding fixed the dismissal propensity.

[40]Predicted recidivism variables are created by regressing recidivism post release if incarcerated, or post conviction/dismissal otherwise, on offense type and sociodemographic controls and month, court, and day-of-the-week fixed effects; then getting the predicted values from the regression. We construct measures of predicted recidivism within one year, within three years, and within five years after case disposition.

6.2 Discussion of the sign and magnitude of the bias

The test above suggests the UPM assumption does not hold in our setting. As discussed in Section 3, when UPM does not hold (but the other assumptions do) 2SLS estimates will be positively weighted averages of the biased Wald estimands derived in Equation 5. Using Equation 5, combined with theory and external evidence, we can reason about the likely sign and magnitude of the bias. We focus our discussion in this section on potential bias in the 2SLS estimates of the impact of conviction relative to dismissal. Throughout this discussion, we will assume that CPM holds, as it does in each of the three models we considered. We also assume A1-A3 and A6 hold.

Equation 5 shows that the bias is proportional to $\Delta_{d\to i}^{Y_i-Y_c} - \Delta_{i\to c}^{Y_i-Y_c}$, or the difference in the impact of incarceration (relative to conviction) between those near the margin with dismissal (those shifted from $d \to i$) and those near the margin with conviction (those shifted from $i \to c$). Hence, we can reason about the likely sign and magnitude of the bias based on conjectures or evidence that inform how incarceration (relative to conviction) may impact recidivism differentially for these two groups.

Table E.10 shows us that the average predicted recidivism rate of the incarcerated group increases in response to increasing Z_c, even as we hold Z_i constant (and therefore the net probability of incarceration constant). This implies that those shifting into incarceration from dismissal have a higher predicted recidivism rate than those shifting out of it to conviction.[41] It's reasonable to think that, in the short run, incarceration affects recidivism primarily through incapacitation (for both groups). If so, shifting prison beds towards those at a higher risk of recidivism will reduce recidivism. In other words, $\Delta_{d\to i}^{Y_i-Y_c} < \Delta_{i\to c}^{Y_i-Y_c}$, hence the bias term in Equation 5 would be negative and our short run estimates would underestimate the increase in recidivism caused by conviction.

If incarceration *only* has incapacitation effects, we would expect the long-run impact of incarceration vs conviction to be zero. This is consistent with the majority of studies on the impact of incarceration (Loeffler and Nagin, 2022), and consistent with RD estimates in our setting, which we discuss in Section 7.2. This would imply that the bias term in Equation 5 is zero, and the long-run estimates of the impact of conviction on recidivism are unbiased.

Of course, it could be the case that incarceration affects recidivism through channels other than incapacitation. For example, prison may be a stronger deterrent after release for people with fewer priors, as in Jordan et al. (2022). Since those with fewer priors are those with lower predicted recidivism, they are overrepresented in the group at

[41] This empirical finding is consistent with a scenario where the individuals on the margin between $T = d$ and $T = i$ are those whose evidence is borderline but the case is serious enough to guarantee incarceration upon conviction, while those on the margin between $T = i$ and $T = c$ have sufficient evidence against them but marginal case severity.

the conviction-incarceration margin, relative to those at the incarceration-dismissal margin. Then, $\Delta_{d\to i}^{Y_i - Y_c} > \Delta_{i\to c}^{Y_i - Y_c}$ and the bias term would be positive. However, we consider this to be unlikely in our setting for two reasons. First, we find no evidence of differential treatment effects of incarceration by prior conviction status (see Panel B of Appendix Table E.5). Second, the RD evidence we present in Section 7.2 shows that incarceration reduces recidivism only in the short run (likely due to incapacitation) for those on the margin of conviction and incarceration. As discussed above, those without priors are likely over-represented for this group, yet we do not find evidence of reduced recidivism beyond the initial period of incapacitation.

Overall, the arguments above suggest that a violation of UPM would lead our 2SLS estimates to have a negative bias in the short run and negligible bias in the long run. Hence, it is unlikely that our qualitative conclusions about the impact of conviction would be overturned as a result of a violation of the UPM assumption.

7 Alternative approaches to identification and estimation of margin-specific treatment effects

This section presents two alternative approaches to identifying and estimating the impacts of conviction and incarceration when treatment effects are heterogeneous. First, in Section 7.1, we discuss an identification approach that is consistent with the unordered multinomial choice model of judge decision-making from Section 5.3. Our approach entails constructing different, treatment-specific instruments, which enable us to recover causal and margin-specific effects. We decompose these causal effects into margin-specific effects using the methods developed in Mountjoy (2022). Second, in Section 7.2, we reconsider the impact of incarceration versus conviction using discontinuities in sentencing guideline recommendations.

7.1 Constructing treatment-specific instruments and decomposing by margin

As described in Section 5.3, under the unordered multinomial model, 2SLS using judge stringency instruments does not identify a causal treatment effect. This section discusses an alternative methodology which, under additional assumptions, allows us to recover the impacts of moving from $T = d$ to $T = c$ and the impacts of moving from $T = i$ to $T = c$. Specifically, we build on the approach to identifying margin-specific treatment effects in Mountjoy (2022). This approach requires treatment-specific instruments, hence we propose constructing such instruments from the panel of judge decisions in our data.

Mountjoy (2022) studies enrollment in two-year and four-year college, using distance to the nearest two-year and four-year colleges as instruments. Unlike stringency instruments, the distance instruments are treatment-specific and shift the cost associated with attending either two-year or four-year college. Varying one distance instrument while holding the other fixed is equivalent to shifting one of the latent thresholds (for example π_c in our setting) while holding the other fixed (for example π_i in our setting).[42] In our notation, this would imply that π_c (the latent threshold for conviction) depends only on z_c (the share of defendants that a judge convicts) and that π_i (the latent threshold for incarceration) depends on only z_i (the share of defendants that a judge incarcerates), which is not the case for our instruments in the unordered multinomial choice setting. This form of variation results in only two groups of movers as visualized in Panel (c) of Figure 5. In Section 7.1.1, we describe our approach for using shares (z_c and z_i) to recover thresholds (π_c and π_i), which we use as treatment-specific instruments.

Even with treatment-specific instruments, the estimands are difficult to interpret, as they would be weighted averages of two LATEs, one for those switching from $T = i$ to $T = c$, and one for those switching from $T = d$ to $T = c$. We refer to the impact of moving from $T = i$ to $T = c$ as "decarceration," as these individuals would no longer be incarcerated. We call the impact of moving from $T = d$ to $T = c$ "labeling," as these individuals would be marked with a felony conviction. The central objective of Mountjoy (2022) is to decompose the 2SLS estimand that mixes these two margin-specific effects into these two separate effects.

7.1.1 Recovering thresholds from choice shares

To apply Mountjoy (2022)'s method, we first conduct an intermediate step of inverting the shares (judge stringencies) we observe for each judge to recover thresholds (π_c and π_i). These thresholds are treatment-specific instruments that allow us to apply the method. We see the shares of cases ending in $T = d$, $T = c$, and $T = i$ for each judge, where individual cases are randomly assigned to each judge. Using the shares, we aim to recover the judge-specific thresholds. We do this by adopting an unordered choice model similar to the model defined in Equation 10, and imposing relatively flexible assumptions about the joint distribution of the error term. Rewriting Equation 10, we

[42]As Mountjoy's method does not use stringency instruments, the implications of UPM are different than in Section 5, when considering 2SLS regressions with judge stringency instruments.

have:

$$\begin{aligned} U_c &= V_c - \pi_c(z_c, z_i) \\ U_i &= V_i - \pi_i(z_c, z_i) \\ U_d &= 0, \end{aligned} \tag{11}$$

where we have normalized the payoff of $T = d$ to zero. The challenge is that π_c and π_i are not known, but we observe the shares of each treatment for each judge. This setup has some similarities to models in industrial organization where shares are observed for different markets, with markets corresponding to judges in our setting.[43] We leverage results from the IO literature and adapt it to our context of judge decision-making. Berry, Gandhi and Haile (2013) show that the inversion between shares and thresholds exists under weak assumptions,[44] and Berry and Haile (2022) implies that judge-specific thresholds can be identified without invoking identification at infinity arguments.[45]

While the π's are identified under relatively weak conditions, we make additional assumptions for tractability in estimation, and show that results are broadly similar under a couple of different sets of assumptions. Our main specification assumes the shocks follow a standard logistic distribution plus a random effect with a correlated multivariate normal distribution. For example, we can then write the payoff to conviction as

$$U_{ncj} = \beta_c + \pi_c^j + \eta_{nc} + \epsilon_{nc},$$

where n represents the case, c indicates this is for conviction, j the judge, and U_{ncj} represents the payoff to a specific outcome for a specific case assigned to judge j.[46] Here we assume $f(\epsilon_{nc}, \epsilon_{ni})$ has a standard logistic distribution and $g(\eta_{nc}, \eta_{ni}) \sim N(0, \Sigma)$. We additionally allow the intercepts and the covariance matrix of the random effect to

[43]Unlike most applications in the IO literature, our setting has quasi-random assignment of cases to judges, implying that $\pi_c(z_c, z_i)$ and $\pi_i(z_c, z_i)$ are independent of V_c and V_i.

[44]To begin, they assume the structural choice probability function can be written with a nonparametric index where judges' latent preferences enter linearly into the index. Then the key assumption is that a "connected substitutes" condition holds. In a multinomial choice setting this condition implies that the probability of choosing j is strictly increasing in the index, which is an input into the structural choice probability function. In a linear-in-parameters unordered choice model, this is satisfied if the support of the additive errors (i.e. the Vs) is $\mathbb{R}^K$, where K is the number of choices.

[45]This proof assumes an index structure on the structural choice probability function where judges' latent preferences enter linearly into the index. Using this setup, the paper shows how the latent judge preferences π^j can be identified using a combination of variation in latent preferences across judges and variation in case characteristics within judge. The proof does not assume the distribution of error terms is independent or identically distributed. Similarly, beyond the assumption on the index function, linearity is not required.

[46]Note that, while we make (flexible) parametric assumptions regarding the joint distribution of V_c and V_i for estimation, we do not make assumptions regarding the relationship between the errors in the choice model and the outcome equations. An alternative approach would be to directly model the joint distribution of error terms in the choice equation and outcomes, such as using a latent factor structure, as in Heckman et al. (2018).

differ across district and by year. Importantly, the random effects allow for correlation between V_c and V_i, and for V_c and V_i to have different variances.[47]

7.1.2 Decomposing effects into labeling and decarceration

We refer to these newly constructed instruments— the estimated judge-specific thresholds —as $\tilde{Z}_c$ and $\tilde{Z}_i$. With these instruments in hand, we closely follow Mountjoy (2022) for estimating the impacts on the two margins discussed above. This method relies on assumptions A1-A3, and A5, defined over $\tilde{z}_c$ and $\tilde{z}_i$. It also requires one additional assumption: 'comparable compliers' (CC). This assumption posits that the $T = i$ to $T = c$ compliers from decreasing $\tilde{z}_i$ have the same potential outcome when convicted as $T = i$ to $T = c$ compliers from increasing $\tilde{z}_c$ at their limits (see Appendix G for a formal definition).

Mountjoy (2022) shows how to identify $E[Y(c) - Y(d)|T = d \to T = c$ complier w.r.t $(\tilde{z}_c, \tilde{z}_i) \to (\tilde{z}_c', \tilde{z}_i)]$ (the labeling effect) and $E[Y(c) - Y(i)|T = i \to T = c$ complier w.r.t $(\tilde{z}_c, \tilde{z}_i) \to (\tilde{z}_c', \tilde{z}_i)]$ (the decarceration effect). The intuition for the approach is that, with the CC assumption, it is possible to construct the expected value of each potential outcome and then to build the treatment effects of interest. We follow Mountjoy (2022) in both the identification and estimation, and provide additional details in Appendix G.

7.1.3 Results

Table 6 shows the results of the decomposition. These results assume a mixed-logit structure with a multivariate normal random effect whose variance and correlation is allowed to vary by court district and year.[48] Panel A reports the labeling effect (C vs D). The results are qualitatively similar to the 2SLS estimates reported in Section 4. Although the point estimates are somewhat smaller than 2SLS, such differences could easily arise by chance given the substantial overlap in the confidence intervals. Panel B reports the decarceration effect (C vs I). Again, results are qualitatively similar to the 2SLS effects of incarceration on recidivism, with the expected change in sign. Panel C reports the net effect: a weighted average of the labeling and decarceration effects. The net effect is statistically significant and positive in the first year (0.037 for future felony charges), and not statistically significant in later years.

[47] In Appendix G, we include additional results which (1) assumes V_c and V_i follow standard logistic distributions and (2) assumes that Σ is a diagonal matrix. Both are less flexible, but easier to implement. For (1), the thresholds are simply $\pi_c(z_c, z_i) = log(z_c) - log(1 - z_c - z_i)$ and $\pi_i(z_c, z_i) = log(z_i) - log(1 - z_c - z_i)$, where the $\mathbf{z}$ are the judge shares.

[48] Tables G.1 and G.2 in Appendix G.2 report results for alternative specifications that assume a standard logit structure and assume the correlation of the random effect is zero, respectively.

Overall, the results from this method tell a similar story to what was seen in the 2SLS estimates. This suggests that any bias induced in 2SLS by a failure of UPM may be small and not qualitatively change the conclusions reached.

7.2 Supporting RD evidence on the impacts of incarceration

Although judges have the final say over sentencing in Virginia, their decisions are influenced by sentencing guidelines. Each person convicted of a felony gets a guidelines-recommended sentence which is calculated using a series of worksheets. Sentence recommendations change discontinuously at some scores. Using a regression discontinuity design, we use this institutional feature to provide additional supporting evidence on the effects of incarceration within the same setting. Exploiting two different discontinuities, we estimate the effects of incarceration on the intensive margin (sentence length) and on the extensive margin (short jail sentences vs probation). We are also able to provide evidence on the extensive margin for those who had never previously been incarcerated. In this section, we provide a brief overview of the empirical framework and main findings, and refer the reader to Appendix H for more details and supporting analyses justifying the empirical approach and expanding on the findings.

7.2.1 Empirical framework

The recommended sentence is calculated based on a series of questions pertaining to the offense and criminal history. Each question has a number of points associated with it, and higher scores lead to tougher sentence recommendations. Based on a person's score, they can be recommended for prison (carceral sentence of a year or more); for jail (carceral sentence under one year); or for probation.

These guidelines present two discontinuous changes in recommendations. We exploit these discontinuities to recover the RD estimates of receiving any incarceration (based on the "probation/jail score") and of sentence length (based on the "sentence-length score"). We use regression discontinuity to compare people who score just below and just above thresholds that lead to more serious sentence recommendations. For these analyses, we use Virginia Criminal Sentencing Commission data, which includes information on all people convicted of a felony in Virginia between 1996 and 2020.[49]

In order for our regression discontinuity research design to be valid, our main assumption is that potential outcomes do not abruptly change at the cutoff.[50] While we

[49]Appendix B.3 describes the sample construction for this analysis.

[50]Since the sentencing scores are discrete, we follow Kolesár and Rothe (2018) to construct "honest" confidence intervals. Appendix H.1 provides details on their approach.

cannot test for this directly, we provide indirect tests as supportive evidence in Appendix H.2, testing for discontinuities in predetermined defendant characteristics (such as demographics or criminal history) and in measures of quality of legal representation at both cutoffs. We generally don't find evidence of discontinuities, suggesting that neither case types nor quality of representation change discontinuously at the cutoffs.

7.2.2 Intensive margin: effects of longer carceral sentences.

As expected from the way worksheets are designed, we find that small differences in the incarceration-length score translate into large changes in people's sentences. Columns 1 and 2 of Table 7 show the regression discontinuity results and Appendix Figure H.4 presents graphic evidence. Scoring above the threshold generates large (42 ppt) changes in the probability of having a sentence greater than one year, and sentences are on average eight months longer, compared to the control-group mean of 4 months.[51] By comparing people on either side of the threshold, we can estimate the causal effect on new criminal justice contact of going from a sentence of approximately four months to approximately one year. Columns 3-9 of Table 7 present outcomes in various time periods, from year 1 to year 8-10 after a person's sentencing date.

Our results are consistent with those estimated using quasi-random assignment of cases to judges. In the first year after sentencing, people above the cutoff are less likely to recidivate. This is likely due to an incapacitation effect: those right below the cutoff have an average sentence of four months, while those right above have an average sentence of 12 months. However, in the longer run, this effect disappears, with no significant difference in recidivism. In our ten year cumulative measure we can reject anything larger than a 1.2 percentage point increase in new felony charges over a control group mean of 46%.

7.2.3 Extensive margin: effects of exposure to incarceration

We found no evidence that tripling the sentence length (from approximately four to 12 months) affected future criminal justice contact. This may be because the impacts of incarceration accrue rapidly in the first several months. For example, a few months in jail might lead a person to lose their job, or to experience ruptures in their family lives (Dobbie et al., 2018). We can test the impact of initial exposure by looking at variation in outcomes for people who score just above or just below the cutoff in the probation/jail score. The first two columns of Panel A of Table 8 show that scoring above the threshold translates into a 43 ppt increase in the likelihood of receiving a carceral sentence, and the average sentence length increases by 0.73 months (Figure H.5

[51]Control-group means are calculated for people whose score is below the relevant cutoff, and whose score is within the bandwidth used in that RD estimate.

presents graphic evidence for this extensive margin). Estimates from the probation/jail sample therefore capture the effect of a short jail sentence relative to probation only.[52] Columns 3-5 of Panel A of Table 8 present results for recidivism. Given that sentences around the cutoff are so short in the sample, we look at short-term results using the six months after sentencing, and longer-term results looking 2-3 years after sentencing. Here, we find no evidence of a short-term incapacitation effect—likely because the difference in sentences is only about a month.[53] As previously, we find no evidence of longer-term effects. In our 1-3 year cumulative measure we can reject anything larger than a 0.007 percentage point increase over a control mean of 20%.

It is also possible that a person's very first incarceration spell may be particularly destabilizing or traumatic. To get at that question, we re-run our analysis on the portion of the probation/jail sample who had not been incarcerated previously, and who had not been detained pretrial.[54] This lowers our sample size substantially, particularly since data on pretrial detention is only available after 2010. As seen in Panel B of Table 8, there is still a strong discontinuity in the likelihood of receiving a carceral sentence for those right above the cutoff, but no evidence of a change in outcomes once the original carceral sentence is complete. However, the estimates are noisy and we can't reject moderate changes in either direction.

These results are very similar to those obtained exploiting quasi-random assignment of cases to judges: we find short-term decreases in criminal justice contact, likely due to incapacitation; but we do not identify any longer-term impacts of exposure to incarceration. Tables E.9 and H.5 present complier characteristics for the IV analyses, and characteristics of defendants who score just above or just below the relevant cutoffs. There are similarities across these groups, but also some small differences. For example, marginal defendants in the RD analysis are more likely to have been convicted with a drug crime compared to the IV compliers—especially for the extensive margin analyses.

8 Conclusion

In this paper, we study the impacts of conviction on future criminal justice contact and compare these estimates to the impacts of incarceration. Across different analyses, we find that conviction increases future criminal justice contact, consistent with a criminogenic effect of a criminal record. In contrast, our analysis of the impact of incarceration only finds evidence for a shorter-term decrease in recidivism, which is

[52]Short sentences such as those experienced right above the cutoff are not atypical. For example, in Pennsylvania, the average amount of time spent in jail post sentencing upon release is 2.4 months (PASC, 2013).

[53]We do find short-term incapacitation effects when looking at quarterly data.

[54]Our data is limited to Virginia; it is possible that they had experienced incarceration in another state.

likely due to incapacitation. Our research shows that criminal justice contact *does* contribute to the revolving door, but through conviction instead of incarceration.

In addition to these substantive findings, our paper presents a discussion of challenges stemming from multiple treatment alternatives in the commonly-used random judge framework. We discuss assumptions that allow us to interpret 2SLS estimates as causal and margin-specific, and which common multiple-outcome choice models are consistent with these assumptions. We propose an approach for testing a key assumption, and derive the asymptotic bias when that assumption does not hold. Based on this expression for the bias, we show that it is possible to reason about its likely sign and magnitude using empirical evidence and features of the institutional setting. Finally, we propose and implement methods to go beyond the 2SLS approach.

These methodological points are relevant in other settings where judges or other 'examiners' choose between multiple alternatives. Multiple alternatives are ubiquitous in the criminal justice setting (Mueller-Smith, 2015; Williams and Weatherburn, 2022; Huttunen et al., 2020; Rivera, 2022), but they also exist in other policy settings where people have used random judge designs, such as foster care (Doyle, 2008; Gross and Baron, 2022; Baron and Gross, 2022), bankruptcy (Dobbie and Song, 2015; Dobbie et al., 2017), or patent decisions (Sampat and Williams, 2019; Feng and Jaravel, 2020; Gavrilova and Juranek, 2021).

Our results relate to several policy debates. Our findings suggest that reducing the scale of incarceration alone is unlikely to shut the revolving door of criminal justice. Rather, policy makers could focus on either reducing the number of people with a felony conviction, or diminishing the channels through which a felony criminal record leads to more recidivism.

Felony convictions could be reduced by increasing felony diversion (Mueller-Smith and Schnepel, 2021; Augustine et al., 2022), decriminalizing certain offenses, or downgrading the charge of conviction to a misdemeanor. Alternatively, the *impact* of felony convictions could be reduced by limiting the accessibility or permissible uses of criminal records. For instance, limiting how long criminal records are publicly available could mitigate employment effects, potentially reducing recidivism by increasing legal employment options (Cullen et al., 2022). Likewise, reducing feedback loops within the penal system, such as automatic charge upgrades or sentence increases for those with a felony conviction, could mitigate concerns that the penal system is itself creating the revolving door problem (Rose, 2021b).

Of course, each of these decisions entail tradeoffs and must take into account a variety of concerns beyond reducing future criminal justice contact. For example, there can be valid reasons for using felony conviction records in the hiring decision or to ratchet up punishment. But, given the prevalence of felony convictions—with

9% of adults and 33% of Black adult men estimated to have a felony criminal record (Shannon et al., 2017)—the impact on future criminal justice contact should be an important part of this discussion.

References

Agan, Amanda Y. and Sonja Starr, "Ban the box, criminal records, and racial discrimination: A field experiment," *The Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

—, Andrew Garin, Dmitri Koustas, Alex Mas, and Crystal S. Yang, "The Impact of Criminal Records on Employment, Earnings, and Tax Outcomes," Working Paper 2022.

—, Jennifer L. Doleac, and Anna Harvey, "Misdemeanor prosecution," Technical Report, National Bureau of Economic Research 2021.

—, Matthew Freedman, and Emily Owens, "Is your lawyer a lemon? Incentives and selection in the public provision of criminal defense," *Review of Economics and Statistics*, 2021, *103* (2), 294–309.

Aizer, Anna and Joseph J. Jr. Doyle, "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges," *Quarterly Journal of Economics*, May 2015, *130* (2), 759–803. MAG ID: 2118051501.

Alper, Mariel, Matthew R. Durose, and Joshua Markman, *2018 update on prisoner recidivism: a 9-year follow-up period (2005-2014)*, US Department of Justice, Office of Justice Programs, Bureau of Justice, 2018.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton university press, 2009.

Arteaga, Carolina, "Parental Incarceration and Children's Educational Attainment," *The Review of Economics and Statistics*, 10 2021, pp. 1–45.

Augustine, Elsa, Johanna Lacoe, Steven Raphael, and Alissa Skog, "The impact of felony diversion in San Francisco," *Journal of Policy Analysis and Management*, 2022, *41* (3), 683–709.

Avi-Itzhak, Benjamin and Reuel Shinnar, "Quantitative models in crime control," *Journal of Criminal Justice*, 1973, *1* (3), 185–217.

Bald, Anthony, Eric Chyn, Justine Hastings, and Margarita Machelett, "The causal impact of removing children from abusive and neglectful homes," *Journal of Political Economy*, 2022, *130* (7), 000–000.

Baron, Jason E. and Max Gross, "Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators," Technical Report, National Bureau of Economic Research 2022.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen, "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections*," *The Quarterly Journal of Economics*, 02 2009, *124* (1), 105–147.

Berry, Steven T., Amit Gandhi, and Philip Haile, "Connected Substitutes and Invertibility of Demand," *Econometrica*, 2013, *81* (5), 2087–2111.

— and **Philip A. Haile**, "Nonparametric Identification of Differentiated Products Demand Using Micro Data," 2022.

Bhuller, Manudeep and Henrik Sigstad, "2SLS with Multiple Treatments," 2022.

— , **Gordon B Dahl, Katrine V. Løken, and Magne Mogstad**, "Incarceration, recidivism, and employment," *Journal of Political Economy*, 2020, *128* (4), 1269–1324.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When is tsls actually late?," Technical Report, National Bureau of Economic Research 2022.

Blevins, Kristie R., Shelley Johnson Listwan, Francis T. Cullen, and Cheryl Lero Jonson, "A general strain theory of prison violence and misconduct: An integrated model of inmate behavior," *Journal of Contemporary Criminal Justice*, 2010, *26* (2), 148–166.

Brayne, Sarah, "Surveillance and system avoidance: Criminal justice contact and institutional attachment," *American Sociological Review*, 2014, *79* (3), 367–391.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk, "Eviction and poverty in American cities: Evidence from Chicago and New York," *Accepted at the Quarterly Journal of Economics*, 2022.

Craigie, Terry-Ann, "Ban the box, convictions, and public employment," *Economic Inquiry*, 2020, *58* (1), 425–445.

Cullen, Zoe B., Will S. Dobbie, and Mitchell Hoffman, "Increasing the Demand for Workers with a Criminal Record," Technical Report, National Bureau of Economic Research 2022.

Deshpande, Manasi and Michael G Mueller-Smith, "Does Welfare Prevent Crime? The Criminal Justice Outcomes of Youth Removed from SSI," Technical Report, National Bureau of Economic Research 2022.

Diamond, Rebecca, Adam Guren, and Rose Tan, "The effect of foreclosures on homeowners, tenants, and landlords," Technical Report, National Bureau of Economic Research 2020.

Dobbie, Will S. and Jae Song, "Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection," *American economic review*, 2015, *105* (3), 1272–1311.

— , **Jacob Goldin, and Crystal S. Yang**, "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 2018, *108* (2), 201–240.

— , **Paul Goldsmith-Pinkham, and Crystal S. Yang**, "Consumer bankruptcy and financial health," *Review of Economics and Statistics*, 2017, *99* (5), 853–869.

Doleac, Jennifer L., "Encouraging Desistance from Crime," *Journal of Economic Literature*, 2023, *61* (2), 383–427.

— and **Benjamin Hansen**, "The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden," *Journal of Labor Economics*, 2020, *38* (2).

Doyle, Joseph J. Jr., "Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care," *Journal of political Economy*, 2008, *116* (4), 746–770.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova, "The deterrent effects of prison: Evidence from a natural experiment," *Journal of political Economy*, 2009, *117* (2), 257–280.

Eichmeyer, Sarah and Jonathan Zhang, "Pathways into Opioid Dependence: Evidence from Practice Variation in Emergency Departments," *American Economic Journal: Applied Economics*, 2022, *14* (4), 271–300.

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai, "Using a probabilistic model to assist merging of large-scale administrative records," *American Political Science Review*, 2019, *113* (2), 353–371.

Estelle, Sarah M. and David C. Phillips, "Smart sentencing guidelines: The effect of marginal policy changes on recidivism," *Journal of public economics*, 2018, *164*, 270–293.

Farrar-Owens, Meredith, "The evolution of sentencing guidelines in Virginia: An example of the importance of standardized and automated felony sentencing data," *Federal Sentencing Reporter*, 2013, *25* (3), 168–170.

Feng, Josh and Xavier Jaravel, "Crafting intellectual property rights: Implications for patent assertion entities, litigation, and innovation," *American Economic Journal: Applied Economics*, 2020, *12* (1), 140–81.

Finlay, Keith, Matthew Gross, Elizabeth Luh, and Michael Mueller-Smith, "The Impact of Financial Sanctions in the U.S. Justice System: Regression Discontinuity Evidence from Michigan's Driver Responsibility Program," Technical Report, Working Paper 2022.

Franco, Catalina, David Harding, Jeffrey Morenoff, and Shawn Bushway, "Failing to Follow the Rules: Can Imprisonment Lead to More Imprisonment Without More Actual Crime," Working Paper 2020.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie, "Judging Judge Fixed Effects," *American Economic Review*, January 2023, *113* (1), 253–277.

French, Eric and Jae Song, "The effect of disability insurance receipt on labor supply," *American economic Journal: economic policy*, 2014, *6* (2), 291–337.

Garin, Andrew, Dmitri Koustas, Carl McPherson, Samuel Norris, Matthew Pecenco, Evan Rose, and Yotam Shem-Tov, "The Impact of Incarceration on Employment and Earnings," Working Paper 2022.

Gavrilova, Evelina and Steffen Juranek, "Female Inventors: The Drivers of the Gender Patenting Gap," *Available at SSRN 3828216*, 2021.

Goldsmith-Pinkham, Paul, Maxim Pinkovskiy, and Jacob Wallace, "The Great Equalizer: Medicare and the Geography of Consumer Financial Strain," *arXiv preprint arXiv:2102.02142*, 2021.

Graef, Lindsay, Sandra G. Mayson, Aurelie Ouss, and Megan T. Stevenson, "Systemic Failures to Appear in Court," *University of Pennsylvania Law Review*, Forthcoming.

Gross, Max and E Jason Baron, "Temporary stays and persistent gains: The causal effects of foster care," *American Economic Journal: Applied Economics*, 2022, *14* (2), 170–99.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman, "The Heavy Costs of High Bail: Evidence from Judge Randomization," *The Journal of Legal Studies*, 2016, *45* (2), 471–505.

Hagan, John, "The social embeddedness of crime and unemployment," *Criminology*, 1993, *31* (4), 465–491.

Harding, David J, Jeffrey D Morenoff, Anh P Nguyen, and Shawn D Bushway, "Imprisonment and labor market outcomes: Evidence from a natural experiment," *American Journal of Sociology*, 2018, *124* (1), 49–110.

Heckman, James J. and Edward J Vytlacil, "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," *Handbook of econometrics*, 2007, *6*, 4875–5143.

_ and Rodrigo Pinto, "Unordered Monotonicity," *Econometrica*, 2018, *86* (1), 1–35.

_ , John Eric Humphries, and Gregory Veramendi, "Dynamic treatment effects," *Journal of Econometrics*, 2016, *191* (2), 276–292. Innovations in Measurement in Economics and Econometrics.

_ , _ , and _ , "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 2018, *126* (S1), S197–S246.

_ , Sergio Urzua, and Edward Vytlacil, "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, 08 2006, *88* (3), 389–432.

_ , _ , and _ , "Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case," *Annals of Economics and Statistics*, 2008, (91/92), 151–174.

Heinesen, Eskil, Christian Hvid, Lars Johannessen Kirkeboen, Edwin Leuven, and Magne Mogstad, "Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study," Technical Report, National Bureau of Economic Research 2022.

Hjalmarsson, Randi, "Juvenile jails: A path to the straight and narrow or to hardened criminality?," *The Journal of Law and Economics*, 2009, *52* (4), 779–809.

Holzer, Harry J, Steven Raphael, and Michael A Stoll, "Perceived criminality, criminal background checks, and the racial hiring practices of employers," *The Journal of Law and Economics*, 2006, *49* (2), 451–480.

_ , _ , and _ , "The effect of an applicant's criminal history on employer hiring decisions and screening practices: Evidence from Los Angeles," *Barriers to reentry*, 2007, *4* (15), 117–150.

Huttunen, Kristiina, Martti Kaila, and Emily Nix, "The Punishment Ladder: Estimating the Impact of Different Punishments on Defendant Outcomes," 2020.

Imbens, Guido and Stefan Wager, "Optimized regression discontinuity designs," *Review of Economics and Statistics*, 2019, *101* (2), 264–278.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

Jordan, Andrew, Ezra Karger, and Derek A Neal, "Heterogeneous Impacts of Sentencing Decisions," *Becker Friedman Institute Working Paper*, 2022, *No. 021-113.*

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad, "Field of Study, Earnings, and Self-Selection," *The Quarterly Journal of Economics*, 2016, *131* (3), 1057–1112.

Kline, Patrick and Christopher R. Walters, "Evaluating Public Programs with Close Substitutes: The Case of Head Start*," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1795–1848.

Kling, Jeffrey R., "Incarceration Length, Employment, and Earnings," *American Economic Review*, June 2006, *96* (3), 863–876.

Kohler-Hausmann, Issa, *Misdemeanorland*, Princeton University Press, 2018.

Kolesár, Michal and Christoph Rothe, "Inference in regression discontinuity designs with a discrete running variable," *American Economic Review*, 2018, *108* (8), 2277–2304.

Kuziemko, Ilyana, "How should inmates be released from prison? An assessment of parole versus fixed-sentence regimes," *The Quarterly Journal of Economics*, 2013, *128* (1), 371–424.

LaCasse, Chantale and A Abigail Payne, "Federal sentencing guidelines and mandatory minimum sentences: Do defendants bargain in the shadow of the judge?," *The Journal of Law and Economics*, 1999, *42* (S1), 245–270.

Lee, Sokbae and Bernard Salanié, "Identifying Effects of Multivalued Treatments," *Econometrica*, 2018, *86* (6), 1939–1963.

Leslie, Emily and Nolan G. Pope, "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments," *The Journal of Law and Economics*, 2017, *60* (3), 529–557.

Lieberman, Carl, Elizabeth Luh, Michael Mueller-Smith, and US CensusBureau UniversityofMichigan UniversityofMichigan, "Criminal court fees, earnings, and expenditures: A multi-state RD analysis of survey and administrative data," Technical Report 2023.

Loeffler, Charles E., "Does Imprisonment Alter the Life Course? Evidence on Crime and Employment From a Natural Experiment," *Criminology*, 2013, *51* (1), 137–166.

— **and Daniel S. Nagin**, "The impact of incarceration on recidivism," *Annual Review of Criminology*, 2022, *5*, 133–152.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand, "Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt," *American economic review*, 2013, *103* (5), 1797–1829.

Mountjoy, Jack, "Community colleges and upward mobility," *American Economic Review*, 2022, *112* (8), 2580–2630.

Mueller-Smith, Michael, "The Criminal and Labor Market Impacts of Incarceration: Identifying Mechanisms and Estimating Household Spillovers," Working Paper August 2015.

— **and Kevin T. Schnepel**, "Diversion in the criminal justice system," *The Review of Economic Studies*, 2021, *88* (2), 883–936.

Natapoff, Alexandra, "Misdemeanors," *S. Cal. L. Rev.*, 2011, *85*, 1313.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver, "The Effects of Parental and Sibling Incarceration: Evidence from Ohio," *American Economic Review*, September 2021, *111* (9), 2926–63.

Pager, Devah, "The mark of a criminal record," *American journal of sociology*, 2003, *108* (5), 937–975.

__ , **Rebecca Goldstein, Helen Ho, and Bruce Western**, "Criminalizing Poverty: The Consequences of Court Fees in a Randomized Experiment," *American Sociological Review*, 2022, *87* (3), 529–553.

PASC, "County Time Served and Revocations: 2013 report," Technical Report, Pennsylvania Commission on Sentencing 2013.

Phelps, Michelle S., "The paradox of probation: Community supervision in the age of mass incarceration," *Law & policy*, 2013, *35* (1-2), 51–80.

__ , "Mass probation: Toward a more robust theory of state variation in punishment," *Punishment & society*, 2017, *19* (1), 53–73.

Philippe, Arnaud, "Learning by doing. How do criminals learn about criminal law?," *University of Bristol, Working Paper*, 2020.

Reaves, Brian A, "Felony defendants in large urban counties, 2009-statistical tables," *Washington, DC: US Department of Justice*, 2013.

Rivera, Roman, "Release, Detain, or Surveil? The Effect of Electronic Monitoring on Defendant Outcomes," Working Paper 2022.

Rose, Evan, "Does Banning the Box Help Ex-Offenders Get Jobs? Evaluating the Effects of a Prominent Example," *Journal of Labor Economics*, 2021, *39* (1).

__ , "Who gets a second chance? Effectiveness and equity in supervision of criminal offenders," *The Quarterly Journal of Economics*, 2021, *136* (2), 1199–1253.

__ **and Yotam Shem-Tov**, "How does incarceration affect reoffending? estimating the dose-response function," *Journal of Political Economy*, 2021, *129* (12), 3302–3356.

Sampat, Bhaven and Heidi L. Williams, "How do patents affect follow-on innovation? Evidence from the human genome," *American Economic Review*, 2019, *109* (1), 203–36.

Shannon, Sarah K. S., Christopher Uggen, Jason Schnittker, Melissa Thompson, Sara Wakefield, and Michael Massoglia, "The Growth, Scope, and Spatial Distribution of People With Felony Records in the United States, 1948-2010," *Demography*, 2017, *54* (5), 1795–1818.

Stevenson, Megan T., "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails," *The Review of Economics and Statistics*, 2017, *99* (5), 824–838.

Sykes, Gresham, *The Society of Captives*, Princeton University Press, 1958.

Vytlacil, Edward, "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 2002, *70* (1), 331–341.

Williams, Jenny and Don Weatherburn, "Can electronic monitoring reduce reoffending?," *The Review of Economics and Statistics*, 2022, *104* (2), 232–245.

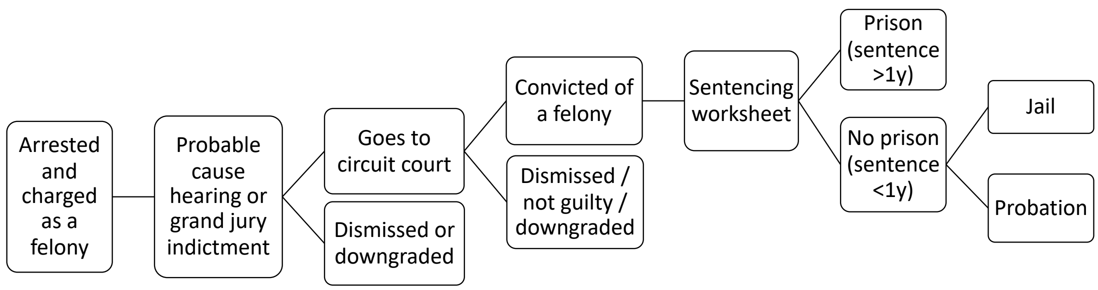
Wolff, Nancy, Cynthia L. Blitz, Jing Shi, Jane Siegel, and Ronet Bachman, "Physical violence inside prisons: Rates of victimization," *Criminal justice and behavior*, 2007, *34* (5), 588–599.

Zimring, Franklin E, Gordon Hawkins, and James Vorenberg, *Deterrence The legal threat in crime control*, University of Chicago Press Chicago, 1973.

9 Figures and tables

9.1 Figures

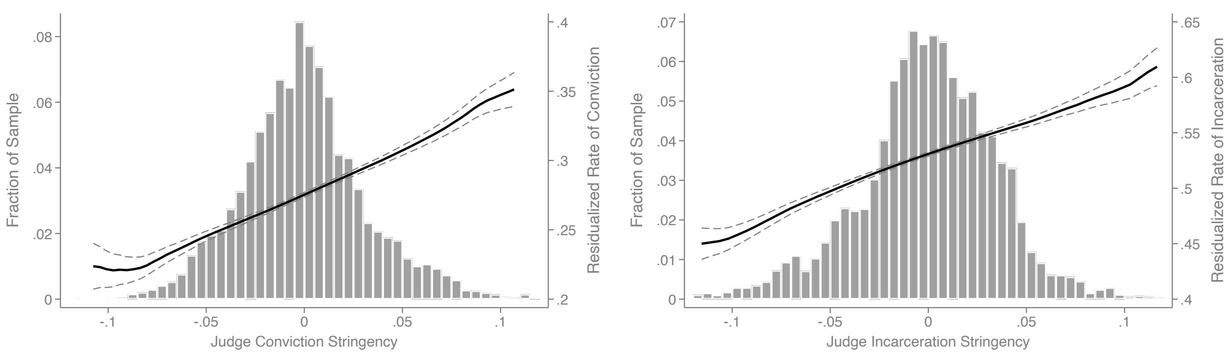
Figure 1: Felony case processing in Virginia, from arrest to disposition



Note: This figure presents a simplified schema of how a felony case flows from arrest to disposition in Virginia.

Figure 2: Distribution of the stringency instruments

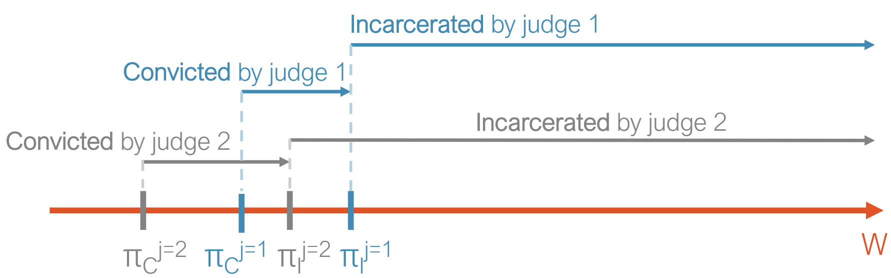
(a) Conviction **(b) Incarceration**



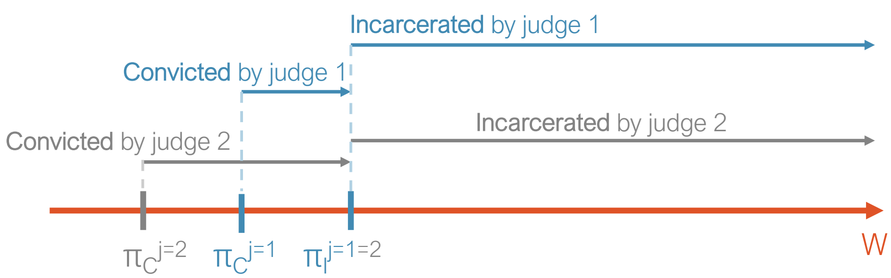
Note: This figure presents our first stages in graphical format for conviction (only), where the outcome is an indicator for the case ending in conviction without incarceration, (Panel A) and incarceration (Panel B). The histograms plot the density of our residualized measures of judge stringency, and the line plots estimates of the first stage regression with conviction (Panel A) and incarceration (Panel B) as the dependent variable.

Figure 3: Ordered choice model

(a)



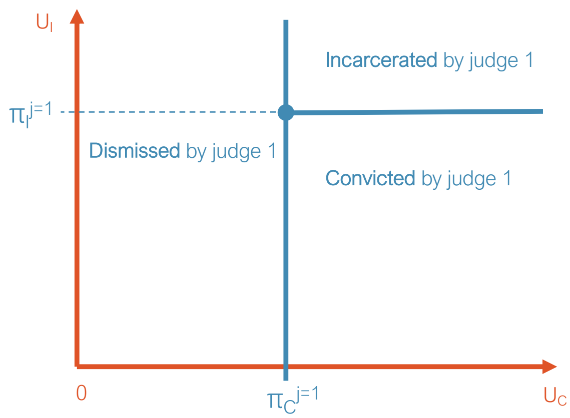
(b)

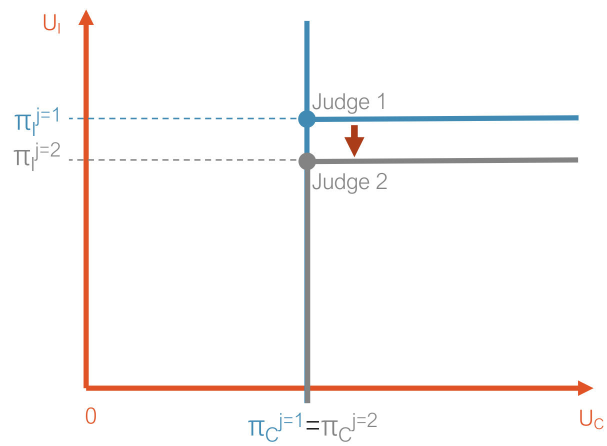


Note: This figure visualizes how, under the ordered choice model discussed in Section 5.1, a judge classifies individuals into incarceration, conviction, and dismissal depending on the cases' unobservable W. Panel A visualizes this for two arbitrary judges, and Panel B does so for two judges with the same incarceration stringency.

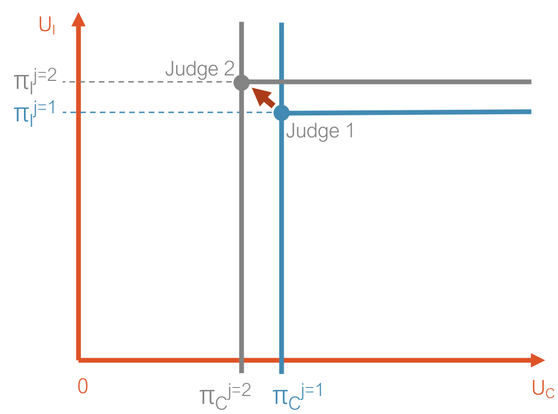
Figure 4: Sequential choice model

(a) A judge divides up (U_c, U_i) *(b) Decreasing* z_i*, holding* z_c *fixed*



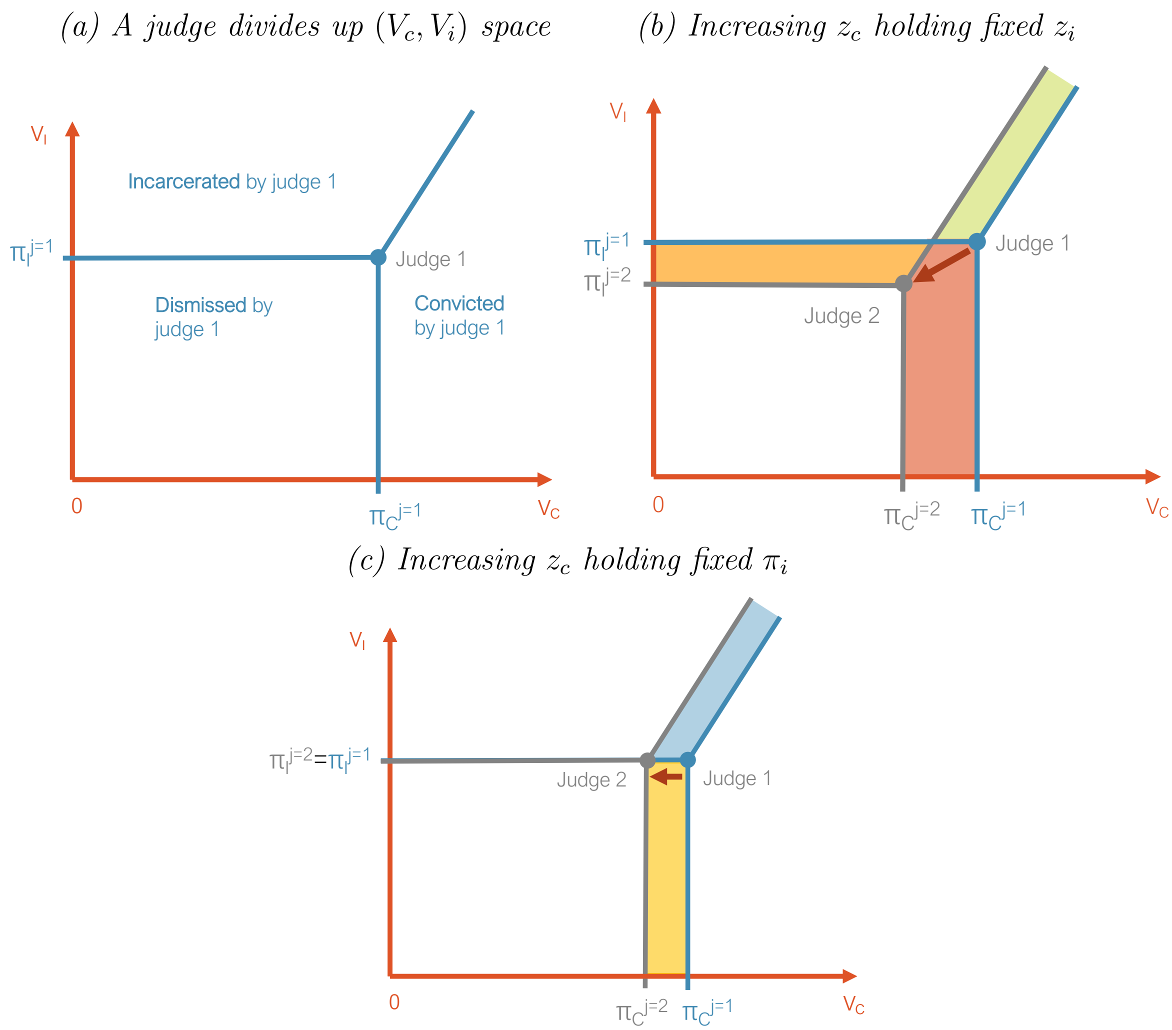


(c) Increasing z_c*, holding* z_i *fixed*



Note: This figure visualizes how, under the sequential choice model discussed in Section 5.2, a judge classifies individuals into incarceration, conviction, and dismissal depending on the cases' unobservable U_i and U_c. Panel A visualizes this for an arbitrary judge, Panel B does so for two judges with the same dismissal stringency and different conviction stringencies, and Panel C for two judges with the same incarceration stringency but where judge 2 has higher conviction stringency.

Figure 5: Unordered multinomial choice model



Note: This figure visualizes how, under the unordered multinomial choice model discussed in Section 5.2, a judge classifies individuals into incarceration, conviction, and dismissal depending on the cases' unobservable V_i and V_c. Panel A visualizes this for an arbitrary judge, Panel B does so for two judges with the same incarceration stringency but where judge 2 has higher conviction stringency, and Panel C for two judges with the same threshold for incarceration but where judge 2 has a higher conviction stringency.

9.2 Tables

Table 1: Summary statistics: 2SLS sample

	Dismissed	Convicted	Incarcerated
	(1)	(2)	(3)
Offenses			
Drugs	0.35	0.34	0.30
Larceny	0.17	0.29	0.26
Assault	0.21	0.08	0.19
Fraud	0.09	0.15	0.10
Traffic	0.03	0.04	0.11
Burglary	0.06	0.06	0.08
Robbery	0.05	0.02	0.06
Sexual assault	0.04	0.02	0.03
Kidnapping	0.03	0.01	0.02
Murder	0.02	0.00	0.01
Defendant characteristics			
Black	0.57	0.51	0.58
Female	0.22	0.32	0.18
% of ppl in zip earning <25K	0.46	0.44	0.46
Incarceration			
Has misdemeanor	0.09	0.11	0.10
prior conviction within 5 years	0.14	0.11	0.23
Incarceration Length	0.00	0.00	26.27
Probation Length	0.00	30.24	37.94
Post-release			
Any charge within 1 year	0.08	0.08	0.06
Median Incar. Leng.	0	0	11
Median Prob. leng.	0	12	12
Percent of Sample	16	29	55
Observations	44,114	79,259	153,692

Note: This table shows means and select medians of relevant variables for the data used in the 2SLS analysis split into the three subsamples. The first column shows estimates for those whose cases were dismissed or who were found not guilty. The second column shows estimates for those whose cases ended with a conviction but without incarceration. The final column shows results for those whose cases ended with incarceration. The summary statistics are for cases closed in 2015 or earlier. The incarceration and probation length medians and means are in months, probation length is top-coded at 20 years.

Table 2: Relevance: first stage coefficients for the 2SLS analysis

	Conviction			Incarceration		
	(1)	(2)	(3)	(4)	(5)	(6)
Conviction Stringency	0.63***	0.60***	0.58***			
	(0.029)	(0.029)	(0.041)			
Incarceration Stringency			-0.020	0.63***	0.61***	0.61***
			(0.036)	(0.029)	(0.028)	(0.031)
Dismissal Stringency						0.016
						(0.045)
Controls	No	Yes	Yes	No	Yes	Yes
Mean Dep. Var.	0.291	0.291	0.291	0.553	0.553	0.553
F-stat	459.8	433.8	197.9	467.7	464.2	385.6
N	231,666	231,666	231,666	231,666	231,666	231,666

Note: This table reports the coefficient on the instruments from the first stage of the 2SLS regressions. Columns (1)-(3) report these coefficients for the conviction analysis, where the outcome is an indicator for the case ending in conviction (without incarceration). The first column includes only the instrument, the second column adds controls about the individual and case, and the third column controls the leave-one-out judge stringency to incarcerate. Column (4)-(6) repeat this analysis, but for the case ending in incarceration, and the final row controlling for the leave-one-out propensity of the judge to end cases with dismissals. Regression includes court-by-year fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. The first stage analysis in this table is on those cases closed in 2015 or earlier. Standard errors are clustered at the judge-year level. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01.

Table 3: Balance

	Convicted	Conv. Stringency	Incarceration	Incar. Stringency
	(1)	(2)	(3)	(4)
Any prior conv.	-0.132***	0.000	0.164***	0.000
	(0.003)	(0.000)	(0.003)	(0.000)
Female	0.114***	-0.000	-0.119***	0.000
	(0.003)	(0.000)	(0.003)	(0.000)
Black	-0.042***	0.000	0.046***	-0.000
	(0.002)	(0.000)	(0.002)	(0.000)
Has misdemeanor	0.039***	0.000	-0.018***	0.000
	(0.004)	(0.000)	(0.004)	(0.000)
Drugs	-0.024***	0.000	0.065***	-0.000
	(0.003)	(0.000)	(0.004)	(0.000)
Larceny	-0.010***	0.000	0.100***	0.000
	(0.003)	(0.000)	(0.003)	(0.000)
Assault	-0.148***	-0.001***	0.147***	0.001***
	(0.003)	(0.000)	(0.004)	(0.000)
Fraud	0.028***	0.000	0.044***	0.001*
	(0.004)	(0.000)	(0.004)	(0.000)
Traffic	-0.186***	-0.000	0.327***	0.001*
	(0.004)	(0.000)	(0.004)	(0.000)
Burglary	-0.045***	-0.000	0.077***	0.000
	(0.004)	(0.000)	(0.004)	(0.000)
Robbery	-0.092***	-0.000	0.155***	0.000
	(0.004)	(0.000)	(0.005)	(0.000)
Sexual assault	-0.168***	-0.001	0.197***	0.001*
	(0.005)	(0.000)	(0.007)	(0.001)
Kidnapping	-0.065***	-0.001	-0.006	0.001
	(0.006)	(0.001)	(0.008)	(0.001)
Murder	-0.148***	-0.001	0.148***	0.001
	(0.007)	(0.001)	(0.011)	(0.001)
P-value joint F-test	0.000	0.000	0.000	0.000
Observations	231,666	231,666	231,666	231,666

Note: This table shows regressions of case characteristics regressed on one of three case outcomes (dismissed, convicted–not incarcerated, or incarcerated). For each outcome, the table reports two sets of regression coefficients. The first regresses an indicator for that outcome on case characteristics, while the second reports the regression of the leave-one-out judge stringency measure for that case outcome. Regression includes court-by-year fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. The offenses are ordered by their prevalence in the data. The balance outcomes shown are for those cases closed in 2015 or earlier. * for $p < .05$, ** for $p < .01$, and *** for $p < .001$.

Table 4: Conviction and recidivism

	Year 1		Year 2-4		Year 5-7		Year 1-4		Year 1-7	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV	(9) OLS	(10) IV
Fut. charge	0.001	0.111***	0.008***	0.049	0.006**	0.077	0.008***	0.140**	0.011***	0.233**
	(0.002)	(0.040)	(0.003)	(0.065)	(0.002)	(0.075)	(0.003)	(0.072)	(0.004)	(0.097)
Fut. conviction	0.003*	0.136***	0.010***	0.071	0.007***	0.054	0.011***	0.205***	0.014***	0.298***
	(0.002)	(0.038)	(0.003)	(0.063)	(0.002)	(0.071)	(0.003)	(0.071)	(0.004)	(0.095)
Fut. incarceration	0.003**	0.109***	0.009***	0.019	0.005**	-0.025	0.010***	0.124**	0.012***	0.214**
	(0.001)	(0.033)	(0.002)	(0.056)	(0.002)	(0.057)	(0.002)	(0.061)	(0.003)	(0.083)
Ctrl Mean: fut. charge	0.069	0.069	0.192	0.192	0.164	0.164	0.239	0.239	0.340	0.340
Ctrl Mean: fut. conv.	0.061	0.061	0.173	0.173	0.149	0.149	0.217	0.217	0.313	0.313
Ctrl Mean: fut. incar.	0.046	0.046	0.135	0.135	0.112	0.112	0.171	0.171	0.250	0.250
Observations	277,065	277,065	231,666	231,666	183,381	183,381	231,666	231,666	183,381	183,381

Note: This table shows regression estimates of the impact of conviction on future recidivism. The five columns report results for five time ranges (1 year, 2-4 years, 5-7 years, 1-4 years, and 1-7 years). For each panel we report ordinary least squares (OLS), and instrumental variable (IV) estimates. Each time period restricts the sample to cases for which the full time period is observed. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. The first three rows report the estimated impact of conviction on different measures of recidivism. The first row is for any future charge, the second row is for any future conviction, and the third row is for any future incarceration. All IV regressions control for the leave-one-out judge incarceration stringency. For the OLS estimates, we regress our measures of recidivism on having a conviction (regardless of incarceration status) controlling for incarceration. The estimates presented are the coefficient on the conviction variable. For the IV estimates, this provides an estimate of the impacts of conviction compared to dismissal for the set of compliers at that margin. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table 5: Incarceration and recidivism

	Year 1		Year 2-4		Year 5-7		Year 1-4		Year 1-7	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV	(9) OLS	(10) IV
Fut. charge	-0.021***	-0.101***	0.012***	-0.010	0.025***	0.004	-0.003	-0.084*	0.023***	-0.070
	(0.001)	(0.024)	(0.002)	(0.040)	(0.002)	(0.040)	(0.002)	(0.045)	(0.003)	(0.059)
Fut. conviction	-0.018***	-0.111***	0.013***	-0.029	0.023***	0.021	-0.001	-0.122***	0.022***	-0.106*
	(0.001)	(0.023)	(0.002)	(0.040)	(0.002)	(0.039)	(0.002)	(0.044)	(0.003)	(0.058)
Fut. incarceration	-0.009***	-0.082***	0.016***	0.012	0.021***	0.053	0.008***	-0.053	0.027***	-0.030
	(0.001)	(0.020)	(0.002)	(0.034)	(0.002)	(0.032)	(0.002)	(0.040)	(0.003)	(0.051)
Ctrl Mean: fut. charge	0.084	0.084	0.171	0.171	0.131	0.131	0.230	0.230	0.303	0.303
Ctrl Mean: fut. conv.	0.074	0.074	0.153	0.153	0.118	0.118	0.207	0.207	0.278	0.278
Ctrl Mean: fut. incar.	0.052	0.052	0.111	0.111	0.084	0.084	0.153	0.153	0.210	0.210
Observations	277,065	277,065	231,666	231,666	183,381	183,381	231,666	231,666	183,381	183,381

Note: This table shows regression estimates of the impact of incarceration on future recidivism. The five columns report results for five time ranges (1 year, 2-4 years, 5-7 years, 1-4 years, and 1-7 years). For each panel we report ordinary least squares (OLS), and instrumental variable (IV) estimates. Each time period restricts the sample to cases for which the full time period is observed. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. The first three rows report the estimated impact of incarceration on different measures of recidivism. The first row is for any future charge, the second row is for any future conviction, and the third row is for any future incarceration. All IV regressions control for the leave-one-out judge dismissal stringency. For the OLS estimates, we regress our measures of recidivism on incarceration, controlling for having a conviction (regardless of incarceration status). For the IV estimates, this provides an estimate of the impacts of incarceration compared to conviction for the set of compliers at that margin. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table 6: Decomposing the impacts of conviction (correlated mixed logit)

	mixed logit with correlated normal random effects				
	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
Panel A: Labeling effect (C vs D)					
Felony Charge:	0.0237	0.103**	0.0897**	0.110	0.140*
	[-0.040,0.102]	[0.001,0.230]	[0.002,0.206]	[-0.019,0.271]	[-0.000,0.284]
Felony Conviction:	0.0429	0.139**	0.0692	0.159***	0.196***
	[-0.024,0.118]	[0.026,0.278]	[-0.037,0.172]	[0.045,0.315]	[0.060,0.330]
Felony Incarceration:	0.0232	0.0811*	0.0289	0.0695	0.111**
	[-0.037,0.087]	[-0.008,0.178]	[-0.047,0.103]	[-0.039,0.204]	[0.004,0.249]
Panel B: Decarceration (C vs I)					
Felony Charge:	0.0404***	-0.0383	-0.00370	-0.0185	0.00135
	[0.014,0.071]	[-0.102,0.017]	[-0.069,0.046]	[-0.092,0.044]	[-0.086,0.084]
Felony Conviction:	0.0342***	-0.0323	-0.0147	-0.0104	-0.00186
	[0.008,0.065]	[-0.086,0.019]	[-0.074,0.048]	[-0.078,0.051]	[-0.094,0.082]
Felony Incarceration:	0.0121	-0.0396*	-0.0209	-0.0280	0.00650
	[-0.012,0.038]	[-0.089,0.002]	[-0.070,0.022]	[-0.089,0.023]	[-0.081,0.079]
Panel C: Net Effect					
Felony Charge:	0.0369***	-0.00788	0.0201	0.00909	0.0367
	[0.013,0.066]	[-0.058,0.039]	[-0.037,0.058]	[-0.040,0.067]	[-0.040,0.103]
Felony Conviction:	0.0360***	0.00458	0.00672	0.0261	0.0486
	[0.014,0.063]	[-0.039,0.048]	[-0.034,0.056]	[-0.024,0.082]	[-0.026,0.116]
Felony Incarceration:	0.0144	-0.0136	-0.00823	-0.00703	0.0332
	[-0.008,0.042]	[-0.053,0.023]	[-0.050,0.023]	[-0.050,0.043]	[-0.041,0.098]
Controls	Yes	Yes	Yes	Yes	Yes

Note: This table decomposes the local IV estimates into the "Labeling effect" (C vs D) and "Decarceration" (C vs I) using an unordered multinomial model, based on the methodology developed in Mountjoy (2022). 95% confidence intervals are reported in brackets and are based on 200 bootstrap samples. The table considers three alternative models for calculating the treatment-specific instrument from the judge's identity. Each panel provides an alternative constructions of the treatment-specific judge instrument. The first panel runs a standard multinomial logit (conditional on the particular court and 3-year bin) that includes judge fixed effects with treatment-specific loadings. Panels B and C run similar specifications but allow the treatment-specific intercepts to include random effects (Panel B) or treatment-specific random effects (Panel C). Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01.

Table 7: Incarceration and recidivism: RD estimates for the intensive margin

	Sentence		Recidivism						
	(1) **Incar > 1 yr**	(2) **Months**	(3) **1 year**	(4) **2-4 years**	(5) **5-7 years**	(6) **8-10 years**	(7) **1-4 years**	(8) **1-7 years**	(9) **1-10 years**
Treatment	0.420	8.005	-0.049	-0.007	0.014	0.002	-0.043	-0.023	-0.016
	[0.404,0.437]	[7.569,8.442]	[-0.065,-0.034]	[-0.024,0.011]	[-0.005,0.034]	[-0.011,0.015]	[-0.065,-0.021]	[-0.048,0.002]	[-0.044,0.012]
N	144,761	144,761	125,876	103,791	81,430	60,350	103,791	81,430	60,350
Control mean	0.08	3.91	0.12	0.23	0.18	0.15	0.31	0.41	0.46

Note: This table first shows the RD estimates of how the cutoff affects sentences (probability of getting a carceral sentence greater than 1 year and sentence length, columns 1 and 2) and recidivism (columns 3-9). We measure recidivism as the likelihood of receiving a new charge for various time windows: the first post-sentencing year, in which incapacitation is most likely, years 2-4, in which some incapacitation may still be present, as well as years 5-7 and years 8-10, during which incarceration rates across treatment and control are equal. It also shows cumulative time windows of 1-4 years, 1-7 years, and 1-10 years to compare to our IV estimates. Below the estimates, we present in brackets confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Appendix H for a discussion of parameter choices.

Table 8: Incarceration and recidivism: RD estimates for the extensive margin

	Sentence		Recidivism		
	(1)	(2)	(3)	(4)	(5)
	Any Incar	**Months**	**6 months**	**2-3 years**	**1-3 years**
Panel A: probation/jail sample					
Treatment:	0.429	0.726	-0.008	-0.003	-0.006
	[0.393,0.465]	[0.532,0.919]	[-0.014,-0.001]	[-0.014,0.009]	[-0.019,0.007]
N	105,839	105,839	91,513	80,320	80,320
Control mean	0.20	0.96	0.06	0.13	0.21
Panel B: no prior incar. probation/jail sample					
Treatment:	0.431	0.820	0.015	-0.000	0.020
	[0.350,0.512]	[0.200,1.440]	[-0.014,0.043]	[-0.043,0.042]	[-0.037,0.076]
N	8,870	8,870	7,871	7,871	7,871
Control mean	0.18	0.81	0.05	0.13	0.20

Note: This table first shows the RD estimates of how the cutoff affects sentences (probability of getting a carceral sentence and sentence length, columns 1 and 2) and recidivism (columns 3-5). We measure recidivism as shows the likelihood of receiving a new charge for various time windows: the first is 6 months post-sentencing year, in which incapacitation is most likely. The second is years 2-3, during which incarceration rates across treatment and control are equal. It also shows cumulative 1-3 year estimates to compare more closely to our IV results. The first panel is our probation/jail score sample while our second panel is for those in our probation/jail sample without prior incarceration post-2010. Below the estimates, we present in brackets confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Appendix H for a discussion of parameter choices.

A Comparing Virginia's criminal justice system to other states

This appendix section shows how Virginia's criminal justice system compares to the U.S. overall, as well as to several states considered in recent related studies: Georgia, Michigan, North Carolina, Ohio, and Texas. First, we re-create figures from Norris et al. (2021) with an additional label for Virginia. Following Norris et al. (2021), we use 2004 data from the Pew Center on three-year recidivism rates, 2004 data on incarceration rates from the Bureau of Justice Statistics, and 2004 data on violent and property crime rates from the FBI Uniform Crime Reporting Program.[55]

Panel (a) of Appendix Figure A.1 shows that while Virginia has similar incarceration rates to the US average and other states, it has slightly lower recidivism (around 28% 3-year recidivism rates). Panel (b) shows that Virginia's property and violent crime rates are lower than the selection of states highlighted, but it is not an outlier in comparison to the rest of the states in the sample.

Appendix Figure A.2 shows prison and jail incarceration rates for the U.S., Virginia, and the five comparison states.[56] Virginia's prison incarceration rate, shown in Panel (a), is 447 per 100,000 people. This is somewhat higher, but comparable to the national rate, and roughly equal to the median among the five comparison states. The rate at which people are jailed in Virginia – 273 per 100,000 – is on the higher end compared to the national average and the five comparison states. Although it is not an obvious outlier relative to either the national average or the five comparison states, when interpreting our results, it is helpful to keep in mind that Virginia tends to rely more on jails than prisons and that conditions may vary across these two settings.

We next consider the racial and ethnic make-up of the prison population in Virginia. Figure A.3 displays the relative ratio of incarceration rates for Black vs White and Hispanic vs White residents.[57] The ratio for Black:White residents in Virginia is 4.3, just below the national average of 4.8 and roughly equal to the average of 4.4 of the other five comparison states. As in others states, Black residents are over-represented in the carceral population. The ratio for Hispanic:White residents is 0.5 for Virginia, lower than national average of 1.3 and most comparison states.

Lastly, we compare probation and parole rates. Virginia's probation rate is close to the national average, as are most comparison states, with the exception of Georgia. However, the parole rate in Virginia – 22 per 100,000 residents – is much lower than the benchmarks. This is because discretionary parole was virtually abolished in Virginia for felonies committed after 1995, resulting in inmates being required to serve at least 85% of their sentences, with the possibility to earn good-time credits toward early release, leading to the small number of individuals on parole in Virginia. This also means that the initial sentence is more closely linked to time spent incarcerated than in other places.

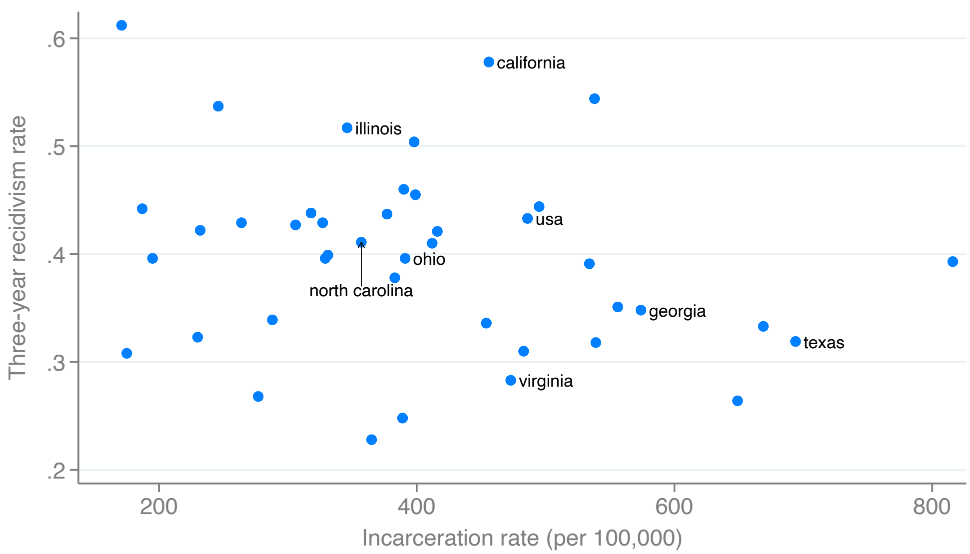
[55] This data can be found at https://www.pewtrusts.org/-/media/legacy/uploadedfiles/pcs_assets/2011/pewstateofrecidivismpdf.pdf, https://bjs.ojp.gov/content/pub/pdf/p04.pdf, and https://www2.fbi.gov/ucr/cius_04/.

[56] We use data from the Prison Policy Initiative. This data can be downloaded from https://www.prisonpolicy.org/reports/correctionalcontrol2018.html.

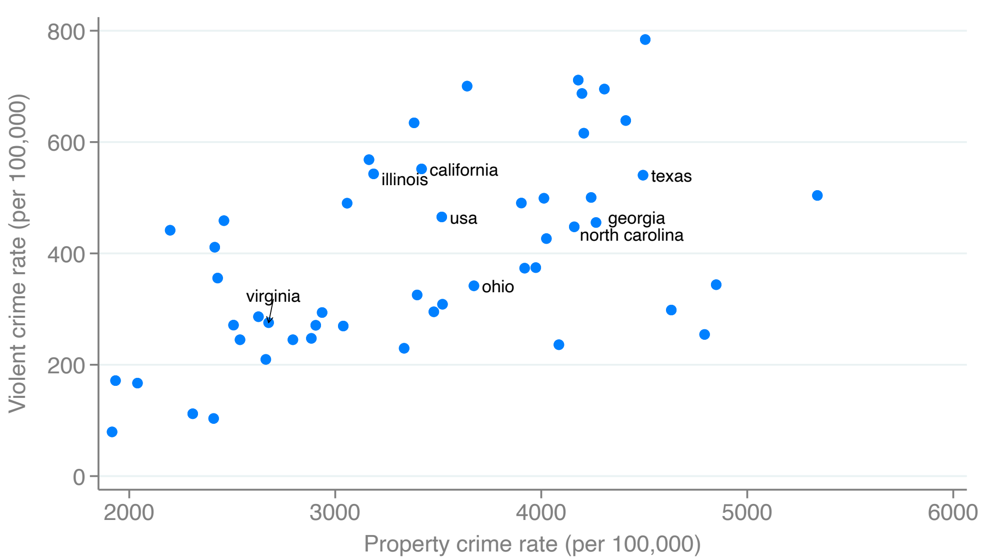
[57] These ratios read as follows: If out of every 100,000 Hispanic residents 200 are incarcerated, and out of every 100,000 White residents 400 are incarcerated, the Hispanic:White ratio is 0.5.

Figure A.1: State-level comparisons of recidivism, incarceration, and crime

(a) 3yr-Recidivism rates vs incarceration rates



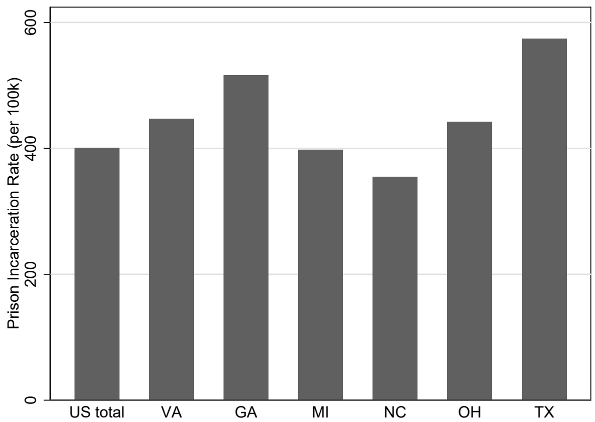
(b) Violent crime rates vs property crime rates



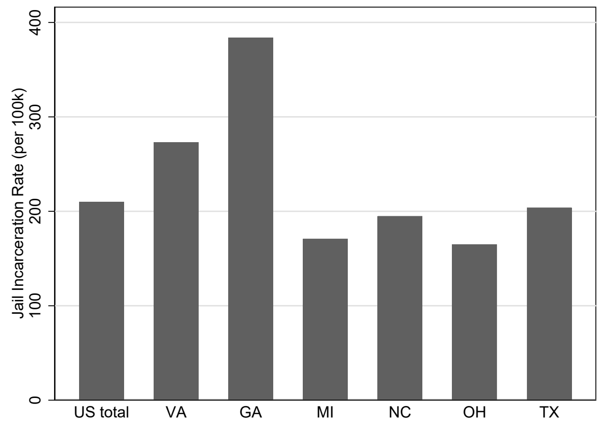
Note: Scatterplots of 2004 incarceration rates, 2004 three-year recidivism rates, and 2004 crime rates. Data gathered from the Pew Center, Bureau of Justice Statistics, and the FBI Uniform Crime Reporting Program.

Figure A.2: Incarceration rates

(a) Prison Incarceration Rates

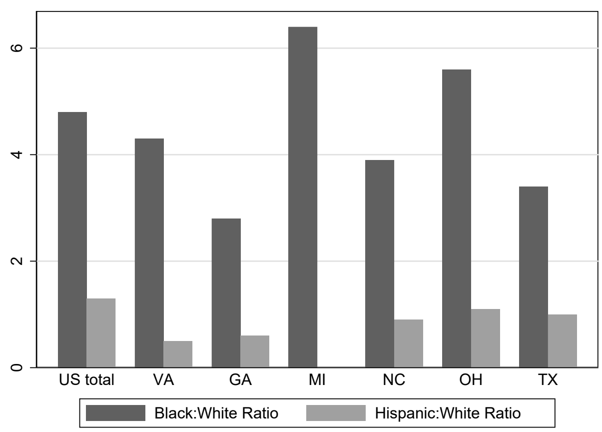


(b) Jail Incarceration Rates



Note: This figure shows the prison (Panel A) and jail (Panel B) incarceration rates, respectively, per 100,000 residents for Virginia, the U.S. overall, Georgia, Michigan, North Carolina, Ohio, and Texas. Based on 2017 and 2014 data respectively from the Prison Policy Initiative (December 2018 press release).

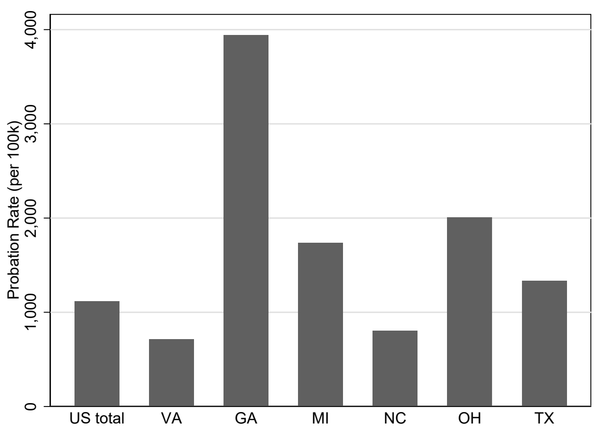
Figure A.3: Racial and ethnic composition of the imprisoned population



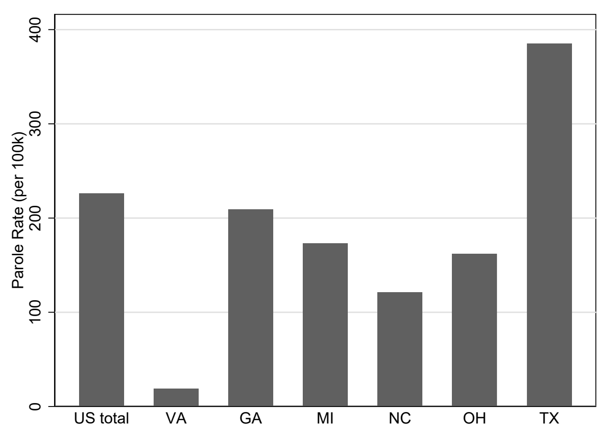
Note: This figure plots the ratio of incarceration rates for Black vs White residents (darker bars) and Hispanic vs White residents (lighter bars), for Virginia, the U.S. overall, Georgia, Michigan, North Carolina, Ohio, and Texas in 2019. Data from sentencingproject.org, used to calculate incarceration by ethnicity, is not available for Michigan.

Figure A.4: Virginia supervision rates comparison

(a) Probation rates



(b) Parole rates



Note: Panel (a) shows the probation rate in Virginia per 100,000 people and Panel (b) shows the parole rate in Virginia per 100,000 people, both compared to the rates for the U.S. total, Georgia, Michigan, North Carolina, Ohio, and Texas. Based on 2016 data from the Prison Policy Initiative (December 2018 press release).

B Additional details on data construction

B.1 Main data sources

Virginia Circuit Courts (VCC) data. The Virginia Court system keeps all Circuit Court case records publicly available for anyone to search. We obtained this data from Ben Schoenfeld who web-scraped records from the courts and made the corresponding data available on http://virginiacourtdata.org/ for public download. This data covers criminal cases in which at least one charge is a felony. It contains information on charges (type and date), on the defendant (gender, race, partial birth date, and FIPS code of residence), and on Circuit Court proceedings for these cases (type, outcome, and judges on the proceedings) and is available for the period 2000-2019. All of Virginia is covered except for Alexandria and Fairfax counties. This is the primary data source for our 2SLS analysis with judge stringencies.

Virginia Criminal Sentencing Commission (VCSC) data. The VCSC provided a dataset that contains information on individuals in Virginia sentenced for a felony. This is used as supplementary data for our 2SLS analysis (to construct our measure of prior convictions), and as the main source for the RD analysis. The data provided to us by the VCSC includes records on all people convicted of a felony in Virginia from 1996 to 2020. This data includes information on the charge(s) of conviction, date of sentencing, sentence imposed for this conviction, guidelines-recommended sentence, points accrued on each item in a worksheet, and total worksheet scores. This data does not contain information on demographics and prior and future charges, so we match it to data from Virginia's Circuit Courts as described below.

B.2 Supplementary data sources

Virginia District Courts (VDC) data. The Virginia Court system also keeps all District Court case records publicly available for anyone to search. As with the Circuit Court data, we obtained this data from Ben Schoenfeld's web-scraped records (http://virginiacourtdata.org/). This data covers all dockets filed in District Court including felonies and misdemeanors. The District Court is a court of limited jurisdiction; felony charges are filed there cannot be adjudicated there. We use this data to obtain information about pretrial detention, as used in the RD specification that subsets to those never previously incarcerated.

Virginia residency data. We obtain information on residency status from a private vendor, matched to the VCSC data with name, social security number and partial birth date. We use the residency data to look at differential mobility in the RD sample. The vendor provided us with information as to which state the matched individual resides in post sentencing. We receive snapshots of information from them 1, 3, 5, 7 years post sentencing date, and we construct a variable indicating if an individual is in the state of Virginia 5 and 7 years post sentencing. 7.7% of observations are missing residency.

IRS ZIP code income data. This is publicly available data produced by the IRS of average ZIP code earnings. We use the 2005 vintage and match in by ZIP onto our samples. This is supplementary data to our IV and RD analysis.

B.3 Data construction

This section details the data construction and cleaning process as well as the matching procedure implemented between the various raw datasets described above.

IV data. We begin with the sample of 3.4 million dockets from the VCC data between 2000 and 2019.

- In addition to dockets with felony charges – the focus of our analysis – the data also includes many dockets pertaining to technical issues (failures to appear in court, revocations, bond hearings, etc.) as well as some pertaining to misdemeanors. We only keep dockets pertaining to new felony charges (roughly 50% of all dockets), leaving roughly 1.6 million felony dockets remaining.[58]
- Sometimes prosecutors file separate dockets for different charges against the same defendant. This could happen if, for instance, the defendant was arrested for multiple burglaries or drug selling occasions. These nonetheless get processed together as one effective case. For our analyses, we define a "case" – our main unit of analysis – as composing all dockets with the same defendant and either the same or consecutive case numbers. Consecutive case numbers means that they were all filed at the same time. Docket level descriptors are aggregated to the case level (i.e. a case is considered 'convicted' if at least one charge was adjudicated guilty). The 1.6 million dockets correspond to 773,553 cases.
- Some courts do not regularly fill out judge information. We drop all courts where less than 80% of judge names are filled out. These courts cover 171,718 cases or 22.2% of cases resulting in 601,835 remaining cases.
- Each case can have multiple hearings. Judge information is provided at the hearing level. We have hearing-level data for 502,732 cases, or 84% of cases.
- We then drop cases entirely missing judge information (37,191 cases dropped or 7.4% of cases resulting in 465,541 cases left).
- We limit ourselves to larger courts with multiple judges overseeing felony cases. In our main sample, we drop judges who see less than 100 cases over 3 years; and all observations in a court-by-year with only one judge. In our main specification, we require that we have at least 3 years per court where multiple judges are present, to avoid including courts and years in which judges simply overlapped because of turnover. In total, these sample restrictions lead us to drop 18,777 cases (4% of the sample), leaving us with 446,764 cases.
- We called clerks in the remaining courts to understand how cases were assigned. In our main specification, we dropped courts where the clerks described a case assignment mechanism that clearly wasn't quasi-random, for instance, ones in which cases are assigned based on judge specialization. We also drop one court only post 2010 due to decreased data availability. This represents 121,931 cases, (27% of remaining cases). This leaves us with 324,799 cases.
- Lastly, we use the VCC data to calculate recidivism, defined as a new felony charge in Circuit Court within X years. The VCC data goes through 2019, and

[58] We also drop dockets that are missing disposition date or initiation date, as well as cases where the disposition is on a weekend. This represents roughly 77,000 dockets, or less than 5% of the remaining sample.

so our analysis sample includes only observations for years before 2019 - X. The largest sample that we use – in which recidivism is defined as new charges within 1 year of conviction – has 277,065 observations.

RD data. We begin by using the VCSC felony data as our universe of cases for each individual convicted of a felony in Virginia. We start with 458,164 observations between 2000 and 2018 (years for which we also have CC data, used to measure recidivism). From there we create two main samples for the RD analyses, as well as a supplementary sample that we use for robustness tests.

Incarceration-length RD data. The first sample leverages the discontinuity in the incarceration-length score as calculated in Worksheet A. This is the sample that we use to measure the effect of longer prison stays vs. shorter jail stays. For this set of analyses, we impose four restrictions on the sample.

- First, we drop offense categories in which the seriousness of the offense mandates a recommended sentence for prison, since we don't have variation at the margins for these cases. The omitted offense categories include murder and voluntary manslaughter, rape, aggravated DWI, some more serious drug offenses (selling to a minor, selling/distributing/manufacturing more heavily restricted drugs/larger quantities), more serious types of assault, burglary, robbery, and other miscellaneous offenses. These constitute roughly 26% of the sample, or 118,364 cases.
- Second, we drop certain offense categories because the distribution of the sentence guidelines scores was very lumpy. Since the RD method requires a smooth evolution of potential outcomes across the running variable, these could be problematic for our design, even if this is mechanically due to the way in which points are accrued. The offense categories dropped are fraud, traffic, and weapons; these constitute 20% of the remaining data, or 72,026 cases. Our main results are robust to including these offense categories.
- Third, we drop individuals who are recorded as having no points in the incarceration-length score: 0.2% of the sample, or 758 cases. We infer that these are likely data errors, since about 10% of these individuals are recommended for prison despite being far below the cutoff at which a prison recommendation is warranted.
- We then match the VCSC sentencing data to the VCC data. VCC data allows us to construct our primary measure of new criminal justice contact (new felony charges in circuit court) as well as race, gender, arrest date, and prior charges. We drop cases from Fairfax and Alexandria, which are not in the CC data. We use the fuzzy matching method developed by Enamorado et al. (2019) and match on first name, last name, middle initial, FIPS code, birth month, and sentence date. For the years and counties in which a match is feasible, our match rate is 92%. Our final sample has 230,357 observations.

Probation/jail RD data. The second sample leverages the discontinuity in the probation/jail score found in Worksheet B. For this set of analyses, we impose similar sample restrictions as described previously.

- First, we drop anyone whose primary offense makes them ineligible for probation, as well as those convicted of violent offenses, since almost none of these are probation-eligible. This represents 59% of the data, or 269,437 cases.
- As previously, we drop individuals who are recorded as having no points in the probation/jail score (0.8%, or 1,576 cases) due to suspected data entry errors. We also drop offense categories for which there are only 2 points between our focal cutoff (probation/jail) and the secondary cutoff (short jail/long jail sentence); which represents 6.8% (or 12,765 cases) of the Worksheet B sample. The remaining offense categories either only have one cutoff (about half of cases) or have 3 points between the focal and secondary cutoff.
- For this data we also restrict to a sample where the Circuit Court match is feasible, using the same procedure as that described for the incarceration-length RD data. Our final sample has 130,692 cases.

Supplementary RD data. Finally, we create a supplementary sample that matches the Worksheet B sample to information on pretrial detention from the VDC data. This reduces our sample significantly since the VDC data is only available from 2010-2019. Since we use three years of follow up, the sample includes those convicted of a felony between 2010-2016: 49,246 cases.

Comparison between IV and RD data. While the data for the RD and the IV analysis comes from the same general sources and have significant overlap there are some key differences.

- The group of cases in the RD data is a subset of those in the 2SLS data, since the RD sample just covers those whose felony charges led to a conviction. For both sets of analyses, we have approximately 80% of Virginia's population since the VCC data misses Alexandria and Fairfax counties.
- In addition, as described above, we further subset the RD sample to include offense types that could, in theory, have led to defendants being on either side of the different RD thresholds.
- Tables 1 and H.1 present summary statistics for each sample.

B.4 Variable construction and definitions

Variable definitions.

- *Incarceration.* We define a person to be incarcerated if at least one of the charges resulted in a positive (greater than zero) carceral sentence.
- *Conviction.* We define a person to be convicted if at least one charge led to a sentence, but no charge resulted in a carceral sentence.
- *Dismissal.* We define a case as dismissed if all charges were dismissed or withdrawn by prosecution (nolle prosequi); or if the defendant was acquitted of all charges.
- *Recidivism.* Our main measure of recidivism is whether a person has a new felony charge in Circuit Court for an offense that allegedly happened after the

focal charge date. This measure does not include revocations, unless these are also accompanied by a new felony charge for a new crime. We create these variables for recidivism in year 1, years 2-4, years 5-7, as well as years 1-4 cumulative and years 1-7 cumulative. For the RD analyses, since we have more years of data, we also include measures for years 8-10 and years 1-10.

- *Recidivism-new conviction.* This is similar to our main recidivism measure, but here the indicator refers to a new conviction on a Circuit Court felony charge for a crime committed within the relevant time periods.
- *Recidivism-new incarceration.* Again, similar to previously, except the indicator means there is a new carceral sentence resulting from a Circuit Court felony charge for a crime committed within the relevant time period.
- *Prior conviction flag.* We define someone as having a prior felony conviction if they have a case in the VCSC data in the 5 years prior to the first offense date of their current case. We use VCSC data to build our prior conviction flag because our data goes back to 1996. This gives us at least 5 years of information on prior felony convictions for all cases in the 2SLS sample.
- *Judge on the case.* We define the judge on the case in the following way. Our main measure is the judge that appears when the "pleading" or the "remarks" variable in the hearings data is marked as "sentencing", "judgement", "dismissal", "conviction", or "final order". If this does not appear on a case, we fill in with the judge present on the disposition date. Finally, if the judge is still missing, for any remaining listings where there is an available judge, we use the maxmode to determine the presiding judge. In our sample, roughly 80% of hearings are in front of the judge whom we define as the judge for the case.[59]
- *Black.* Race of the defendant as defined in the VCC data. Almost all of the people for which race information is available are labeled either "Black" or "White." Ethnicity is not available.
- *Female.* Gender of the defendant as defined in the VCC data.
- *Incarceration Length.* This variable indicates how long in months an individual is imprisoned (if they have a carceral sentence). It will be 0 otherwise.
- *Income generating.* This is a variable that is used to determine whether the individual has new felony charges for an income-generating type of crime. We consider the following charges to be income-generating:burglary, drug charges (excluding drug possession), fraud, larceny, robbery, or prostitution.
- *Has misdemeanor.* An indicator if the current case has a misdemeanor charge as recorded in the Circuit Court data.
- *Under 23.* An indicator defined if the defendant is under 23 year of age of offense. This comes from VCSC data.
- *% of people in ZIP earning <25K.* Share of people earning less than 25K in a ZIP code, using matched IRS average ZIP code level earnings data.

[59] The other hearings could be seen by another judge because the primary judge is absent that day (sick or on vacation) or if the case was reassigned.

C Additional details on bias in 2SLS estimands

C.1 The Wald derivation under CPM and UPM

Proposition 1 in the main paper shows that, under CPM, the Wald estimand from comparing two judges with different conviction stringencies but the same incarceration stringency is a weighted average of LATEs for a specific margin, plus an additive bias term that must be weakly smaller in magnitude than the difference in LATEs for a different margin for two equally-sized groups of compliers.

Consider increasing conviction stringency from z_c to z_c' while holding incarceration stringency z_i fixed.[60] Let $\omega_{i\to c}$ represent the proportion of people switching from $T=i$ to $T=c$. To keep notation simple, we suppress notation indicating the values of the instruments, for example, we write $\omega_{d\to c}$ rather than $\omega_{d\to c}(z_c', z_c|z_i)$. Similarly, allow $\omega_{d\to c}$ and $\omega_{c\to i}$ represent the proportions of people in the other two sets of potential switchers. Next, let $\Delta_{i\to c}^{Y_c-Y_i}$ represent the local average Y_c-Y_i treatment effect for those who switched from $T=i$ to $T=c$ when the instrument shifted from z_c to z_c', holding z_i fixed. More generally, let $\Delta_{j\to k}^{Y_m-Y_n}$ be the treatment effect of moving from $T=n$ to $T=m$ for j to k compliers when the instrument shifted from z_c to z_c', holding z_i fixed.

When CPM holds but UPM does not, a movement from z_c to z_c' holding z_i fixed induces three types of flows: $d\to c$, $d\to i$, and $i\to c$. The reduced form effect is thus given by:

$$E[Y(z_c', z_i) - Y(z_c, z_i)] = \omega_{d\to c}\Delta_{d\to c}^{Y_c-Y_d} + \omega_{i\to c}\Delta_{i\to c}^{Y_c-Y_i} + \omega_{d\to i}\Delta_{d\to i}^{Y_i-Y_d}. \tag{1}$$

The derivation of the bias rests on two observations. First, since the overall probability of incarceration is fixed at z_i, the proportion of the sample flowing into and out of incarceration must be equal in size (i.e., $\omega_{d\to i} = \omega_{i\to c}$). Second, the treatment effects from the unwanted flows, e.g. the flows from $d\to i$, can be decomposed into desired treatment effects ($d\to c$) and undesired ones ($c\to i$). Thus, Equation 1 can be rewritten as:

$$\begin{aligned} E[Y(z_c', z_i) - Y(z_c, z_i)] &= \omega_{d\to c}\Delta_{d\to c}^{Y_c-Y_d} + \omega_{i\to c}\left[\Delta_{i\to c}^{Y_c-Y_i} + \Delta_{d\to i}^{Y_i-Y_c} + \Delta_{d\to i}^{Y_c-Y_d}\right] \\ &= \omega_{d\to c}\Delta_{d\to c}^{Y_c-Y_d} + \omega_{i\to c}\Delta_{d\to i}^{Y_c-Y_d} + \omega_{i\to c}\left[\Delta_{d\to i}^{Y_i-Y_c} - \Delta_{i\to c}^{Y_i-Y_c}\right]. \end{aligned}$$

Similarly, $E[T_c(z_c', z_i) - T_c(z_c, z_i)] = (\omega_{i\to c} + \omega_{d\to c})$. Constructing the Wald estimand we then have:

$$\frac{E[Y(z_c', z_i) - Y(z_c, z_i)]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} =$$

$$\underbrace{\frac{\omega_{d\to c}\Delta_{d\to c}^{Y_c-Y_d} + \omega_{i\to c}\Delta_{d\to i}^{Y_c-Y_d}}{\omega_{d\to c} + \omega_{i\to c}}}_{\text{Weighted avg. of } Y_c - Y_i \text{ treatment effects}} + \underbrace{\frac{\omega_{i\to c}}{\omega_{d\to c} + \omega_{i\to c}}\left[\Delta_{d\to i}^{Y_i-Y_c} - \Delta_{i\to c}^{Y_i-Y_c}\right]}_{\text{Bias term}}. \tag{2}$$

The first term is the weighted average two LATEs, both of moving from $T=d$ to $T=c$,

[60] We derive the bias for impacts of conviction vs dismissal here; the bias for other margins is analogous and only requires rearranging subscripts.

but for two different groups of compliers. The second term is the bias and its formula illuminates two facts. First, the bias is proportional to the fraction of compliers that comes from the undesired margin: $i \to c$ instead of $d \to c$. Second, the sign of the bias depends on the difference in the effects of incarceration ($\Delta^{Y_i - Y_c}$) between the two subgroups: those induced from $d \to i$ by the change in conviction stringency, and those induced from $i \to c$. For example, if the treatment effect is smaller for the former than the latter, then the estimates will be negatively biased, since $\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c}$ would be less than zero. Similarly, if the treatment effect $Y(i) - Y(c)$ is constant, the bias term is zero.

Moving from CPM to the stronger UPM assumption further simplifies Equation (2). First, recall that UPM($Z_c|Z_i$) implies that there can only be flows into $T = c$ when increasing Z_c from z_c to z_c'. Second, recall that fixing judge stringeny $Z_i = z_i$ implies that the net probability of incarceration must remain constant. This second point implies that any flows from $T = i$ to $T = c$ would need to be compensated by flows from $T = d$ to $T = i$. Since UPM($Z_c|Z_i$) rules out flow from $T = d$ to $T = i$, there can be no flows from $T = i$ to $T = c$ since Z_i is fixed. This implies that $\omega_{i \to c} = 0$, which simplifies Equation 2 to

$$\frac{E[Y(z_c', z_i) - Y(z_c, z_i)]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} = \Delta_{d \to c}^{Y_c - Y_d}.$$

C.2 Bias with four treatments

Here, we calculate the bias from a 2SLS estimate in a simple setting with four mutually exclusive treatments. For example, these could be dismissed; convicted without incarceration; convicted with a short carceral sentence; or convicted with a long carceral sentence: $T \in \{d, c, s, l\}$. The mutually-exclusive stringencies would then be: Z_d, Z_c, Z_s, Z_l. We assume CPM and the other assumptions, except for UPM (see Section 3.1 for details).

In the example below, we characterize bias when using differential stringencies to determine the causal effect of conviction vs dismissal. Let's consider two judges who have the same z_s and z_l, but different z_c. Following the notation from Appendix C.1, ω's represent shares of switchers. For example, $\omega_{d \to c}$ represents the proportion of people switching from $T = d$ to $T = c$ when shifting conviction stringency from z_c to z_c', holding z_s and z_l fixed.

The set of potential movers when changing z_c (holding fixed z_s and z_l) under CPM are: (1) $d \to c$, (2) $s \to c$, (3) $l \to c$, (4) $d \to s$, (5) $d \to l$, and (6) $l \to s$. Note that this is just one possible direction of switches that would be compatible with CPM. For instance, for (6), we could have reversed flows, and allowed for $s \to l$ instead $l \to s$; but under CPM we can only have one, not the other. The same applies for (5).

As with 3 treatments, holding z_s fixed means that flows in and out of $T = s$ have to be equal, and holding z_l fixed means flows in and out of $T = l$ have to be equal. This means that $\omega_{s \to c} = \omega_{l \to s} + \omega_{d \to s}$ and $\omega_{d \to l} = \omega_{l \to s} + \omega_{l \to c}$.

The reduced form effect is thus given by:

$$E[Y(z_c', z_s, z_l) - Y(z_c, z_s, z_l)] = \tag{3}$$
$$\Delta_{d \to c}^{Y_c - Y_d}\omega_{d \to c} + [\Delta_{d \to s}^{Y_s - Y_d}\omega_{d \to s} + \Delta_{s \to c}^{Y_c - Y_s}\omega_{s \to c}] + [\Delta_{d \to l}^{Y_l - Y_d}\omega_{d \to l} + \Delta_{l \to c}^{Y_c - Y_l}\omega_{l \to c}] + \Delta_{l \to s}^{Y_s - Y_l}\omega_{l \to s},$$

where brackets have been placed around two sets of terms to simplify the explanation of the next steps below.

For any difference in two potential outcomes, we can always rewrite it as $Y_k - Y_j = (Y_k - Y_m) - (Y_j - Y_m)$. Using this, the first term in the square brackets in Equation 3 can be rewritten as follows:

$$\begin{aligned}
[\Delta^{Y_s-Y_d}_{d\to s}\omega_{d\to s} + \Delta^{Y_c-Y_s}_{s\to c}\omega_{s\to c}] &= [\Delta^{Y_s-Y_d}_{d\to s}\omega_{d\to s} + \Delta^{Y_c-Y_s}_{s\to c}(\omega_{d\to s} + \omega_{l\to s})] \\
&= [\Delta^{Y_s-Y_d}_{d\to s}\omega_{d\to s} + (\Delta^{Y_c-Y_d}_{s\to c} - \Delta^{Y_s-Y_d}_{s\to c})\omega_{d\to s} + \Delta^{Y_c-Y_s}_{s\to c}\omega_{l\to s}] \\
&= [\Delta^{Y_c-Y_d}_{s\to c}\omega_{d\to s} + (\Delta^{Y_s-Y_d}_{d\to s} - \Delta^{Y_s-Y_d}_{s\to c})\omega_{d\to s} + \Delta^{Y_c-Y_s}_{s\to c}\omega_{l\to s}].
\end{aligned} \tag{4}$$

Similarly, the second term in the square brackets from Equation 3 can be rewritten:

$$\begin{aligned}
[\Delta^{Y_l-Y_d}_{d\to l}\omega_{d\to l} + \Delta^{Y_c-Y_l}_{l\to c}\omega_{l\to c}] &= [\Delta^{Y_l-Y_d}_{d\to l}(\omega_{l\to s} + \omega_{l\to c}) + \Delta^{Y_c-Y_l}_{l\to c}\omega_{l\to c}] \\
&= [\Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to s} + \Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to c} + (\Delta^{Y_c-Y_d}_{l\to c} - \Delta^{Y_l-Y_d}_{l\to c})\omega_{l\to c}] \\
&= [\Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to s} + \Delta^{Y_c-Y_d}_{l\to c}\omega_{l\to c} + (\Delta^{Y_l-Y_d}_{d\to l} - \Delta^{Y_l-Y_d}_{l\to c})\omega_{l\to c})].
\end{aligned} \tag{5}$$

So, Equation 3 can be written as:

$$\begin{aligned}
E[Y(z'_c, z_s, z_l) - Y(z_c, z_s, z_l)] = \\
&\Delta^{Y_c-Y_d}_{d\to c}\omega_{d\to c} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{d\to s} + \Delta^{Y_c-Y_d}_{l\to c}\omega_{l\to c} \\
&+ (\Delta^{Y_s-Y_d}_{d\to s} - \Delta^{Y_s-Y_d}_{s\to c})\omega_{d\to s} \\
&+ (\Delta^{Y_l-Y_d}_{d\to l} - \Delta^{Y_l-Y_d}_{l\to c})\omega_{l\to c} \\
&+ \Delta^{Y_s-Y_l}_{l\to s}\omega_{l\to s} + \Delta^{Y_c-Y_s}_{s\to c}\omega_{l\to s} + \Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to s}.
\end{aligned} \tag{6}$$

Next, the last row of Equation 6 can be rewritten as:

$$\begin{aligned}
&\Delta^{Y_s-Y_l}_{l\to s}\omega_{l\to s} + \Delta^{Y_c-Y_s}_{s\to c}\omega_{l\to s} + \Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to s} \\
&= \Delta^{Y_s-Y_l}_{l\to s}\omega_{l\to s} + (\Delta^{Y_c-Y_d}_{s\to c}\omega_{l\to s} - \Delta^{Y_s-Y_d}_{s\to c}\omega_{l\to s}) + \Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to s} \\
&= \Delta^{Y_s-Y_l}_{l\to s}\omega_{l\to s} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{l\to s} - (\Delta^{Y_s-Y_l}_{s\to c}\omega_{l\to s} + \Delta^{Y_l-Y_d}_{s\to c}\omega_{l\to s}) + \Delta^{Y_l-Y_d}_{d\to l}\omega_{l\to s} \\
&= \Delta^{Y_c-Y_d}_{s\to c}\omega_{l\to s} + (\Delta^{Y_s-Y_l}_{l\to s} - \Delta^{Y_s-Y_l}_{s\to c})\omega_{l\to s} + (\Delta^{Y_l-Y_d}_{d\to l} - \Delta^{Y_l-Y_d}_{s\to c})\omega_{l\to s}.
\end{aligned} \tag{7}$$

Rewriting Equation 6, we get:

$$\begin{aligned}
E[Y(z'_c, z_s, z_l) - Y(z_c, z_s, z_l)] = \\
&\Delta^{Y_c-Y_d}_{d\to c}\omega_{d\to c} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{d\to s} + \Delta^{Y_c-Y_d}_{l\to c}\omega_{l\to c} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{l\to s} \\
&+ (\Delta^{Y_s-Y_d}_{d\to s} - \Delta^{Y_s-Y_d}_{s\to c})\omega_{d\to s} \\
&+ (\Delta^{Y_l-Y_d}_{d\to l} - \Delta^{Y_l-Y_d}_{l\to c})\omega_{l\to c} \\
&+ (\Delta^{Y_s-Y_l}_{l\to s} - \Delta^{Y_s-Y_l}_{s\to c})\omega_{l\to s} + (\Delta^{Y_l-Y_d}_{d\to l} - \Delta^{Y_l-Y_d}_{s\to c})\omega_{l\to s}.
\end{aligned} \tag{8}$$

And the first row of Equation 8 can be rewritten as:

$$\begin{aligned}
&\Delta^{Y_c-Y_d}_{d\to c}\omega_{d\to c} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{d\to s} + \Delta^{Y_c-Y_d}_{l\to c}\omega_{l\to c} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{l\to s} \\
=&\Delta^{Y_c-Y_d}_{d\to c}\omega_{d\to c} + \Delta^{Y_c-Y_d}_{s\to c}(\omega_{d\to s} + \omega_{l\to s}) + \Delta^{Y_c-Y_d}_{l\to c}\omega_{l\to c} \\
=&\Delta^{Y_c-Y_d}_{d\to c}\omega_{d\to c} + \Delta^{Y_c-Y_d}_{s\to c}\omega_{s\to c} + \Delta^{Y_c-Y_d}_{l\to c}\omega_{l\to c}.
\end{aligned} \tag{9}$$

Equation 3 can thus be expressed in terms of $d \to c$ treatment effects (first line of Equation 10) and differences in the same treatment effects between different subgroups (remaining lines of Equation 10):

$$E[Y(z_c', z_s, z_l) - Y(z_c, z_s, z_l)] = \tag{10}$$
$$\underbrace{\Delta^{Y_c - Y_d}_{d \to c}\omega_{d\to c} + \Delta^{Y_c - Y_d}_{s \to c}\omega_{s\to c} + \Delta^{Y_c - Y_d}_{l \to c}\omega_{l\to c}}_{\text{Weighted } d \to c \text{ treatment effects}}$$
$$\underbrace{\begin{aligned} &+ (\Delta^{Y_s - Y_d}_{d\to s} - \Delta^{Y_s - Y_d}_{s\to c})\omega_{d\to s} \\ &+ (\Delta^{Y_l - Y_d}_{d\to l} - \Delta^{Y_l - Y_d}_{l\to c})\omega_{l\to c} \\ &+ (\Delta^{Y_s - Y_l}_{l\to s} - \Delta^{Y_s - Y_l}_{s\to c})\omega_{l\to s} \\ &+ (\Delta^{Y_l - Y_d}_{d\to l} - \Delta^{Y_l - Y_d}_{s\to c})\omega_{l\to s} \end{aligned}}_{\text{Differences in subgroup treatment effects}}$$

Next, the denominator of the Wald estimator will be given by:

$$E[T_C(z_c', z_s, z_l) - T_C(z_c, z_s, z_l)] = \omega_{d\to c} + \omega_{s\to c} + \omega_{l\to c}, \tag{11}$$

Finally, dividing Equation 10 by Equation 11, we end up with two terms. The first term is a weighted average of margin-specific treatment effects of moving from $T = d$ to $T = c$ for three groups of compliers. The weights here are all weakly positive and sum to one. The second term is a weighted average of the four bias terms, where each term is the difference in the treatment effect of a given margin for two different sets of compliers, and the weights are weakly positive. This implies that the bias will depend on the heterogeneity of treatment effects. For example, under a constant effects assumption, the bias terms are all zero.

Note that this expression parallels the expression derived in Appendix C.1 where we have a proper weighted average of the margin-specific effects of interest and an additive weighted bias term, where the size of the bias depends on how heterogeneous the margin-specific treatment effects are.

C.3 Interpreting conditional 2SLS estimates

In the main paper, we consider the comparison of two judges that have the same stringency on one margin, but different stringencies on another margin. For example, for the Wald estimands, we consider two judges that have the same incarceration stringency $Z_i = z_i$, but different conviction stringencies Z_c. Here, we consider what the IV estimand identifies when exclusion, random assignment, relevance, and the conditional pairwise monotonicity (CPM) assumptions hold, and how this changes when swapping out CPM for the unordered partial monotonicity assumption (UPM). Specifically, we consider the case where we first condition on a set of judges who have the same incarceration stringency $Z_i = z_i$ but potentially differ in their conviction stringency. We assume Z_c can take on values $\{z_c^0, ..., z_c^K\}$ where the set is ordered such that $z_c^k \leq z_c^{k'}$ if $k \leq k'$.

In Appendix C.1, we derive the Wald estimand when comparing two judges with the same incarceration stringency but different conviction stringencies. This gives us:

$$\text{Wald}(z_c', z_c|z_i) = \frac{E[Y(z_c', z_i) - Y(z_c, z_i)]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} = \frac{E[Y|Z_c = z_c', Z_i = z_i] - E[Y|Z_c = z_c, Z_i = z_i]}{E[T_c|Z_c = z_c', Z_i = z_i] - E[T_c|Z_c = z_c, Z_i = z_i]} =$$

$$\underbrace{\frac{\omega_{d\to c}\Delta_{d\to c}^{Y_c - Y_d} + \omega_{i\to c}\Delta_{d\to i}^{Y_c - Y_d}}{\omega_{d\to c} + \omega_{i\to c}}}_{\text{Weighted avg. of } Y_c - Y_i \text{ treatment effects}} + \underbrace{\frac{\omega_{i\to c}}{\omega_{d\to c} + \omega_{i\to c}}\left[\Delta_{d\to i}^{Y_i - Y_c} - \Delta_{i\to c}^{Y_i - Y_c}\right]}_{\text{Bias term}}.$$

Now, we derive what is identified in this setting by IV when using judges with varying conviction stringency but the same incarceration stringency. For notational simplicity, we leave the conditioning on $Z_i = z_i$ implicit throughout this derivation. The IV estimand is given by: $\alpha^{IV} = \frac{E[Y(Z_c - E[Z_c])}{E[T_c(Z_c - E[Z_c]]} = \frac{cov(Y, Z_c)}{cov(T_c, Z_c)}$ Following Imbens and Angrist (1994) closely, first consider the numerator:

$$\begin{aligned}
E[Y \cdot (Z_c - E[Z_c])] &= \sum_{l=0}^{K} \lambda_l E[Y|Z_c = z_c^l](z_c^l - E[Z_c]) \\
&= \sum_{l=0}^{K} \lambda_l E[Y|Z_c = z_c^0](z_c^l - E[Z_c]) \\
&+ \sum_{l=1}^{K} \lambda_l \sum_{k=1}^{l} \left(E[Y|Z_c = z_c^k] - E[Y|Z_c = z_c^{k-1}]\right)(z_c^l - E[Z_c]) \\
&= \sum_{k=1}^{K} \left(\left(E[Y|Z_c = z_c^k] - E[Y|Z_c = z_c^{k-1}]\right) \sum_{l=k}^{K} \lambda_l (z_c^l - E[Z_c])\right) \\
&= \sum_{k=1}^{K} \text{Wald}(z_c^k, z_c^{k-1}|z_i) \left(\left(E[T_c|Z_c = z_c^k] - E[T_c|Z_c = z_c^{k-1}]\right) \sum_{l=k}^{K} \lambda_l (z_c^l - E[Z_c])\right)
\end{aligned}$$

Next, the denominator using a similar set of steps can be written as:

$$\begin{aligned}
E[T_c(Z_c - E[Z_c])] &= \sum_{l=0}^{K} \lambda_l E[T_c|Z_c = z_c^l](z_c^l - E[Z_c]) \\
&= \sum_{k=1}^{K} \left(\left(E[T_c|Z_c = Z_c^k] - E[T_c|Z_c = Z_c^{k-1}]\right) \sum_{l=k}^{K} \lambda_l (z_c^l - E[Z_c])\right)
\end{aligned}$$

Putting these together, we get:

$$\alpha^{IV} = \sum_{k=1}^{K} \theta_{k,k-1} \text{Wald}(z_c^k, z_c^{k-1}|z_i)$$

where

$$\theta_{k,k-1} = \frac{\left(E[T_c|Z_c = z_c^k] - E[T_c|Z_c = z_c^{k-1}]\right)\sum_{l=k}^{K}\lambda_l(z_c^l - E[Z_c]))}{\sum_{k=1}^{K}\left(E[T_c|Z_c = z_c^k] - E[T_c|Z_c = z_c^{k-1}]\right)\sum_{l=k}^{K}\lambda_l(z_c^l - E[Z_c])}.$$

Other than the implicit conditioning on $Z_i = z_i$, this formula is the same as the formula derived in Imbens and Angrist (1994), but the Wald estimand may not always be a pairwise LATE as in Imbens and Angrist (1994). Under the CPM assumption and other standard IV assumptions, the Wald estimand recovers the term given in Equation 5 in Section 3. Thus, rather than a weighted average of pairwise local-average treatment effects, we recover a weighted average of the potentially-biased margin-specific local average treatment effects. Under the stronger UPM assumption, or under a constant-effects assumption, Equation 5 reduces down to a standard margin-specific LATE as in Imbens and Angrist (1994) and the conditional 2SLS estimand can be interpreted as a positively weighted average of LATEs where the weights sum to one.

Based on these results, a natural path forward would be to estimate separate 2SLS regressions, conditional on each value of Z_i. Angrist and Pischke (2009) propose doing this in a single 2SLS regression where the instrument Z_c is interacted with all possible values of Z_i. They refer to this as the "saturate and weight" approach. However, in finite samples, this approach can result in many weak instruments and the problems that arise in such setting (Angrist and Pischke, 2009; Blandhol et al., 2022).

Tables C.1 and C.2 show estimates where the treatment and instrument have been interacted with the other judge stringency. Some caution in interpreting these estimates as splitting our sample even into thirds quickly leads to large standard errors and small first-stage F-statistics. We report four specifications which include increasingly rich sets of controls which are described in the table notes. Across all specifications the majority of estimates are positive, nearly all estimates are positive when including richer controls, and all negative estimates are statistically insignificant with very large standard errors. Across estimates we see very similar trends with large increases in conviction in the first year that accumulate over time.

Table C.1: **The impacts of conviction on recidivism: interacting treatment and instruments with incarceration stringency bins**

	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
	(1)	(2)	(3)	(4)	(5)
Specification 1:					
Convict x bottom 3rd	0.130**	-0.051	0.097	0.043	0.142
	(0.056)	(0.110)	(0.096)	(0.116)	(0.160)
Convict x middle 3rd	0.265***	0.240	0.267	0.394*	0.609**
	(0.095)	(0.208)	(0.166)	(0.207)	(0.290)
Convict x top 3rd	0.369	-0.754	0.218	-0.460	-0.249
	(0.296)	(0.628)	(0.500)	(0.615)	(0.860)
Specification 2:					
Convict x bottom 3rd	0.139**	-0.015	0.151	0.081	0.207
	(0.068)	(0.129)	(0.141)	(0.145)	(0.222)
Convict x middle 3rd	0.316***	0.196	0.397	0.397	0.731*
	(0.118)	(0.220)	(0.249)	(0.243)	(0.403)
Convict x top 3rd	0.352	-0.638	0.349	-0.361	-0.053
	(0.308)	(0.593)	(0.620)	(0.633)	(0.953)
Specification 3:					
Convict x bottom 3rd	0.116**	0.075	0.119	0.171**	0.285**
	(0.047)	(0.074)	(0.089)	(0.086)	(0.124)
Convict x middle 3rd	0.150**	0.191*	0.233*	0.274**	0.472***
	(0.064)	(0.108)	(0.120)	(0.121)	(0.172)
Convict x top 3rd	0.184**	-0.016	0.212	0.111	0.290
	(0.075)	(0.129)	(0.162)	(0.151)	(0.244)
Specification 4:					
Convict x bottom 3rd	0.063***	-0.005	0.002	0.038	0.035
	(0.024)	(0.038)	(0.038)	(0.044)	(0.054)
Convict x middle 3rd	0.085***	0.043	0.007	0.094**	0.060
	(0.029)	(0.040)	(0.045)	(0.045)	(0.062)
Convict x top 3rd	0.097***	0.040	0.091**	0.068	0.112*
	(0.027)	(0.043)	(0.042)	(0.048)	(0.060)
Observations	277,065	231,666	183,381	231,666	183,381

Note: This figure shows 2SLS estimates of the impact of conviction on future charges. Each specification interacts conviction and conviction stringency with residualized incarceration stringency terciles. Specification 1 includes our standard set of fixed effects: court-by-year, court-by-month of year, and day-of-week dummies. Specification 2 replaces court-by-year and court-by-month of year dummies with court-by-year-by-month of year dummies. As the tercile interactions only condition on three bins of incarceration stringency, Specification 3 further adds dummies for deciles of residualized judge incarceration stringency. The final specification replaces the conviction instrument interacted with residualized incarceration stringency terciles with judge dummies. Standard errors are clustered at the judge-year level. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01.

Table C.2: The impacts of incarceration on recidivism: interacting treatment and instruments with conviction stringency bins

	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
	(1)	(2)	(3)	(4)	(5)
Specification 1:					
Incarceration x bottom 3rd	-0.066	-0.078	0.012	-0.164	-0.128
	(0.059)	(0.092)	(0.095)	(0.101)	(0.128)
Incarceration x middle 3rd	-0.163***	0.011	-0.038	-0.106	-0.092
	(0.048)	(0.069)	(0.083)	(0.081)	(0.123)
Incarceration x top 3rd	-0.012	0.023	0.055	0.047	0.102
	(0.048)	(0.084)	(0.076)	(0.095)	(0.115)
Specification 2:					
Incarceration x bottom 3rd	-0.072	-0.092	-0.015	-0.189*	-0.166
	(0.064)	(0.096)	(0.101)	(0.107)	(0.135)
Incarceration x middle 3rd	-0.173***	0.010	-0.031	-0.115	-0.106
	(0.051)	(0.075)	(0.088)	(0.087)	(0.133)
Incarceration x top 3rd	0.009	0.048	0.067	0.082	0.139
	(0.052)	(0.088)	(0.080)	(0.101)	(0.122)
Specification 3:					
Incarceration x bottom 3rd	-0.219	-0.052	0.019	-0.206	-0.147
	(0.276)	(0.111)	(0.079)	(0.141)	(0.110)
Incarceration x middle 3rd	-0.371	0.026	-0.045	-0.215	-0.171
	(0.492)	(0.180)	(0.168)	(0.237)	(0.243)
Incarceration x top 3rd	0.435	-0.009	-0.004	0.174	0.090
	(1.070)	(0.295)	(0.150)	(0.400)	(0.226)
Specification 4:					
Incarceration x bottom 3rd	-0.104***	-0.050	-0.051	-0.115***	-0.129***
	(0.018)	(0.031)	(0.032)	(0.034)	(0.043)
Incarceration x middle 3rd	-0.087***	-0.047	-0.042	-0.110***	-0.109**
	(0.021)	(0.033)	(0.036)	(0.037)	(0.051)
Incarceration x top 3rd	-0.052***	0.047	0.072**	0.034	0.142***
	(0.020)	(0.033)	(0.035)	(0.038)	(0.051)
Observations	277,065	231,666	183,381	231,666	183,381

Note: This figure shows 2SLS estimates of the impact of incarceration on future charges. Each specification interacts incarceration and incarceration stringency with residualized dismissal stringency terciles. Specification 1 includes our standard set of fixed effects: court-by-year, court-by-month of year, and day-of-week dummies. Specification 2 replaces court-by-year and court-by-month of year dummies with court-by-year-by-month of year dummies. As the tercile interactions only condition on three bins of dismissal stringency, Specification 3 further adds dummies for deciles of residualized judge dismissal stringency. The final specification replaces the incarceration instrument interacted with residualized dismissal stringency terciles with judge dummies. Standard errors are clustered at the judge-year level. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01.

C.4 Average UPM

$\text{UPM}_{Z_c|Z_i}$ represents a form of "strict" monotonicity, in that it is defined over every z_c shift, holding z_i constant. Yet, similar to what has been shown in the binary context, such a strict assumption is not necessary to yield a causal estimand. Frandsen et al. (2023) propose a condition called "average monotonicity," which requires a positive correlation between each individual's *potential* treatment status and judge stringency across all judges. They show that average monotonicity is sufficient (along with other standard IV assumptions) to yield a causal estimand in the binary treatment context.

Here we propose an extension of Frandsen et al. (2023)'s average monotonicity condition into the three treatment setting and refer to this as "average $\text{UPM}_{Z_c|Z_i}$." We focus on the condition that is relevant to the specification where we are instrumenting for conviction and controlling for the incarceration stringency; average $\text{UPM}_{Z_i|Z_d}$ is defined similarly.

We first introduce an additional piece of notation. Let G be a group variable where $g \in G$ maps $(Z_c,\ Z_i)$ onto potential treatment $T_c(Z_c, Z_i)$. G is the collective and mutually exclusive set of groups g. In the binary treatment, binary instrument context, G consists of compliers, defiers, always takers, and never takers.

A5b: Average $\text{UPM}_{Z_c|Z_i}$: For all (g, z_i) in the support of (G, Z_i) the following conditions must hold:

$$Cov(T_c(Z_c, Z_i), Z_c|Z_i = z_i, G = g) \geq 0 \tag{12}$$

$$Cov(T_i(Z_c, Z_i), Z_c|Z_i = z_i, G = g) = 0 \tag{13}$$

To illustrate a difference between $\text{UPM}_{Z_c|Z_i}$ and average $\text{UPM}_{Z_c|Z_i}$, consider a shift from z_c to $z'_c > z_c$, holding z_i constant. If there exists a group g for whom this instrument shift would induce them from conviction to dismissal, $\text{UPM}_{Z_c|Z_i}$ would be violated but average $\text{UPM}_{Z_c|Z_i}$ might not be. As long as the probability of conviction for each group is positively correlated with the overall conviction propensity of judges, average $\text{UPM}_{Z_c|Z_i}$ is satisfied.

Average $\text{UPM}_{Z_c|Z_i}$, along with A1-A3 and A6, is sufficient for Equations 3 and 4 to yield margin-specific and causal estimands. We build off of Blandhol et al. (2022) for the proof. First, note that the second line of A5b, combined with A2 and A3 (random assignment and exclusion) assure that the exogeneity condition outlined in Blandhol et al. (2022) is met. In our setting, this exogeneity condition means that $G, Y(T = c) \perp Z_c|Z_i$. G is orthogonal to Z_c (conditional on Z_i) due to the random assignment assumption. $Y(T = c)$ is orthogonal to Z_c because, if you hold Z_i fixed, Z_c will not be correlated with the probability of incarceration for any group.

With exogeneity in hand, the remainder of the proof is provided by Blandhol et al. (2022). Blandhol et al. (2022) focus on a condition they call "monotonicity-correct," which they show is sufficient for the 2SLS estimator with controls to be weakly causal (i.e. the weights on all group-specific treatment effects are weakly positive and the estimate does not depend on the levels of the dependent variable). In the appendix, they derive the monotonicity condition that is both sufficient and necessary for weakly causal estimates, which is the condition in line one of A5b, when written in our notation and in the terms relevant to our setting.[61] They do not focus on this condition in the main text because "such fortuitous averaging would be difficult to defend." In the judge IV context, however, this "fortuitous averaging" could naturally occur. For instance, a judge who punishes harshly overall may be relatively lenient on certain types of offenders. This would violate both the monotonicity-correct condition as well as UPM. But as long as relatively harsh judges increase punishment *on average* for all groups, an occasional judge who bucks the trend for certain groups is not a problem.

C.5 Interpreting 2SLS estimates with controls

Appendix section C.3 derived the 2SLS estimand when conditioning on a specific value of Z_i. The estimation results reported in Section 4 control for Z_i rather than condition. This section discusses how to interpret these 2SLS estimates. In particular, following Blandhol et al. (2022), 2SLS specifications that control for Z_i (and potentially other covariates) can still be interpreted as a positively weighted sum of the Wald estimates we derived in Section 3, as long as one additional assumption is met.

Blandhol et al. (2022) considers what 2SLS recovers when covariates are included as controls, but not fully saturated as in the "saturate and weight" approach. They show that covariates can introduce substantial bias and result in estimands that are not what they call "weakly causal." They define an estimand as weakly causal when it (i) does not depend on the levels of the potential outcomes when holding treatment effects (differences) constant and (ii) it does not apply negative weights to any subgroup. Blandhol et al. (2022) goes on to discuss what assumptions are necessary and sufficient for 2SLS with controls to recover weakly causal parameters. For our setting, with a scalar multi-valued instrument, one additional assumption needs to hold:[62]

A6b. Rich covariates: the linear projection of Z on X is equal to the conditional expectation of Z given X. That is $L[Z|X] = X'E[XX']^{-1}E[XZ] = E[Z|X]$.

Assumption A6b implies that we need to include a rich set of controls. Note that assumption A6b differs from assumption A6 as Section 3.4 abstracted away from covariates. Here we provide the more general version of the assumption which allows for other covariates. When the only covariate is Z_i, this implies we need rich controls for Z_i. When instruments are only randomly assigned conditional on a vector of covariates $\mathbf{X}$, then we must include a sufficiently rich set of controls for the full vector of covariates, including Z_i.

Blandhol et al. (2022)'s Proposition 11 provides an expression for what the 2SLS estimand recovers. A small rearrangement of that expression allows it to be written as a positively weighted average of Wald estimands. Under assumptions A1-A4 or A1-A3 and A5, these Wald estimands are equivalent to those we derive in Sections

[61] The sufficient and necessary condition for weakly causal estimates is presented in the paragraph between Equation 28 and Equation 29 in the appendix proof for Proposition 9 (page 50) of the version from August 9, 2022. Our Z_c would be written $\dot{Z}$ in their notation, our Z_i would be their X, and our $T_c^g(Z_c)$ would be $\mathbb{1}(Z \in \mathbb{Z}_j(g))$.

[62] Note that assumption A1-A3, and A5 satisfy the other needed assumptions in Blandhol et al. (2022). In particular, A5 implies their "Ordered strong monotonicity" (OSM). Assumption A4 also satisfies the OSM, but violates their definition of exclusion, which can result in biased Wald estimates, similar to those we derive under CPM.

3.3 and 3.4. Thus, under assumptions A1-A4, and A6b, 2SLS recovers a positively weighted average of terms which are margin-specific causal effects plus additive bias terms. Under assumptions A1-A3, A5, and A6b, 2SLS recovers a positively weighted average of margin-specific treatment effects.

Tables C.3 and C.4 show that our estimates are not sensitive to the richness of our control variables. Each specification adds increasingly detailed sets of dummies for place, time, and the other judge stringency as described in the table notes. All specifications are similar to the estimates we report in the main paper, and trend towards larger estimates when including richer set of controls. Like our main estimates we find large increases in recidivism from conviction that accumulate over time, while incarceration has a negative effect in the first year which remains relatively constant when looking at one year, one to four year, or one to seven years.

Table C.3: The impacts of conviction on recidivism: robustness to richness of controls

	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
	(1)	(2)	(3)	(4)	(5)
Specification 1:					
Fut. charge	0.128***	0.099	0.105	0.201**	0.298***
	(0.044)	(0.072)	(0.079)	(0.080)	(0.108)
Fut. conviction	0.151***	0.131*	0.074	0.275***	0.365***
	(0.042)	(0.069)	(0.075)	(0.080)	(0.106)
Fut. incarceration	0.119***	0.055	-0.007	0.167**	0.254***
	(0.036)	(0.060)	(0.060)	(0.069)	(0.093)
Specification 2:					
Fut. charge	0.132***	0.131	0.155	0.233**	0.361***
	(0.051)	(0.081)	(0.094)	(0.093)	(0.129)
Fut. conviction	0.157***	0.164**	0.122	0.311***	0.437***
	(0.048)	(0.078)	(0.091)	(0.094)	(0.127)
Fut. incarceration	0.127***	0.073	0.032	0.194**	0.327***
	(0.041)	(0.069)	(0.073)	(0.079)	(0.112)
Fut. incarceration	0.076*	0.024	-0.057	0.120	0.202*
	(0.043)	(0.071)	(0.069)	(0.081)	(0.116)
Specification 3:					
Fut. charge	0.151***	0.116	0.097	0.217**	0.325***
	(0.048)	(0.079)	(0.082)	(0.088)	(0.116)
Fut. conviction	0.180***	0.148**	0.069	0.303***	0.408***
	(0.047)	(0.075)	(0.077)	(0.089)	(0.116)
Fut. incarceration	0.142***	0.064	-0.027	0.182**	0.280***
	(0.040)	(0.066)	(0.062)	(0.075)	(0.101)
Specification 4:					
Fut. charge	0.163***	0.151*	0.149	0.259**	0.396***
	(0.058)	(0.090)	(0.100)	(0.106)	(0.143)
Fut. conviction	0.196***	0.179**	0.118	0.347***	0.491***
	(0.056)	(0.087)	(0.096)	(0.109)	(0.145)
Fut. incarceration	0.157***	0.074	0.009	0.208**	0.355***
	(0.047)	(0.076)	(0.076)	(0.089)	(0.124)
Observations	277,065	231,666	183,381	231,666	183,381

Note: This table reports estimates of the impact of conviction on our three measures of recidivism. Each specification adds richer controls. Specification 1 includes the fixed effects included in the paper: court-by-year, court-by-month of year, and day of week dummies, plus percentile dummies for residualized judge incarceration stringenecy. Specification 2 matches specification 1 but swaps out court-by-year and court-by-month of year fixed effects with court-by-year-by-month of year fixed effects. Specification 3 includes the main place and location fixed effects plus year-by-decile of residualized incarceration stringency dummies. Specification 4 is the same as specification three, but swaps out year-by-court and year-by-month of year dummies with court-by-year-by-month of year dummies. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table C.4: The impacts of incarceration on recidivism: robustness to richness of controls

	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
	(1)	(2)	(3)	(4)	(5)
Specification 1:					
Fut. charge	-0.102***	-0.034	-0.008	-0.110**	-0.093
	(0.024)	(0.039)	(0.040)	(0.044)	(0.059)
Fut. conviction	-0.112***	-0.051	0.007	-0.144***	-0.128**
	(0.023)	(0.039)	(0.039)	(0.044)	(0.058)
Fut. incarceration	-0.083***	-0.007	0.043	-0.074*	-0.049
	(0.020)	(0.033)	(0.032)	(0.039)	(0.050)
Specification 2:					
Fut. charge	-0.099***	-0.033	-0.013	-0.110**	-0.100
	(0.025)	(0.041)	(0.043)	(0.047)	(0.063)
Fut. conviction	-0.112***	-0.052	0.006	-0.146***	-0.135**
	(0.024)	(0.041)	(0.042)	(0.046)	(0.062)
Fut. incarceration	-0.083***	-0.013	0.039	-0.080*	-0.063
	(0.021)	(0.035)	(0.034)	(0.041)	(0.054)
Specification 3:					
Fut. charge	-0.103***	-0.033	0.002	-0.115**	-0.085
	(0.026)	(0.039)	(0.041)	(0.045)	(0.060)
Fut. conviction	-0.115***	-0.048	0.009	-0.146***	-0.122**
	(0.025)	(0.038)	(0.040)	(0.044)	(0.059)
Fut. incarceration	-0.085***	-0.010	0.034	-0.080**	-0.059
	(0.022)	(0.033)	(0.033)	(0.040)	(0.053)
Specification 4:					
Fut. charge	-0.103***	-0.042	-0.015	-0.125***	-0.108*
	(0.028)	(0.042)	(0.045)	(0.048)	(0.064)
Fut. conviction	-0.117***	-0.059	-0.004	-0.160***	-0.146**
	(0.027)	(0.041)	(0.043)	(0.048)	(0.064)
Fut. incarceration	-0.087***	-0.029	0.024	-0.100**	-0.090
	(0.024)	(0.036)	(0.035)	(0.043)	(0.057)
Observations	277,065	231,666	183,381	231,666	183,381

Note: This table reports estimates of the impact of incarceration on our three measures of recidivism. Each specification adds richer controls. Specification 1 includes the fixed effects included in the paper: court-by-year, court-by-month of year, and day of week dummies, plus percentile dummies for residualized judge dismissal stringenecy. Specification 2 matches specification 1 but swaps out court-by-year and court-by-month of year fixed effects with court-by-year-by-month of year fixed effects. Specification 3 includes the main place and location fixed effects plus year-by-decile of residualized dismissal stringency dummies. Specification 4 is the same as specification three, but swaps out year-by-court and year-by-month of year dummies with court-by-year-by-month of year dummies. Standard errors are clustered at the judge-year level. Stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Validating assumptions A1-A4

In this section, we discuss whether Assumptions A1-A4 from Section 3 are supported by features of the institutional environment and provide empirical evidence, based on a standard battery of tests, to help assess their validity.

Relevance. Here, we explain the various ways through which judges can influence both conviction and incarceration outcomes. We also present empirical evidence of the relevance of judges' influence on these decisions.

Judges influence conviction in several ways. In all cases, they have the latitude to dismiss charges if they find the evidence insufficient. They are directly responsible for adjudicating guilt during bench trials (that is, trials by judge, without lay jurors). They also exert indirect influence on the likelihood of conviction via several channels. First, they make the determination on various pretrial motions, which can have a large impact on the likelihood of conviction. For example, they can refuse to grant a continuance if a key witness does not show up to court on a given day. They rule on the admissibility of evidence, including critical pieces like confessions, possession of contraband, or expert testimony. Finally, they can affect jury composition by ruling on motions to strike, and by formulating jury instructions.

Judges also influence sentences in several ways. In the case of a bench trial, they directly choose the sentence. In the case of guilty pleas, they can reject the negotiated plea agreement. Their reputation as a tough or a lenient judge might shape what offers prosecutors and defense attorneys are willing to put forward (LaCasse and Payne, 1999). Similarly, if the judge has a reputation for choosing short sentences, the prosecutor may adjust and offer shorter sentences as part of the plea deal.

Empirically, we find persistent differences in case outcomes across judges. Figure 2 in the main paper shows the histogram of judge conviction (without incarceration) stringency in panel (a) and judge incarceration stringency in panel (b). Each panel plots the residualized leave-one-out judge propensity for that case outcome. In both panels there is substantial variation in the instrument.[63] Both panels also plot the local linear regression of the residualized court outcome on the instrument.

Appendix Figure D.1 plots the residualized conviction and incarceration stringencies against each other. The two instruments are negatively correlated, which is expected, since the probability of all three case outcomes adds up to one. Importantly for our research design, there is substantial variation in z_c across most of the support of z_i and vice versa.

Table 2 in the main paper presents our first-stage estimates, and confirms that judge stringency has a large and statistically significant effect on conviction and incarceration. The first three columns show the results for the first stage on conviction. The first column shows the loading on conviction stringency when only including interacted court and time fixed effects as controls. The second column adds detailed case-level controls. The third column additionally controls for incarceration stringency. Across all three specifications, the conviction stringency remains large, with partial F-statistics between 198 and 460. Columns 4 through 6 perform similar first-stage regressions on

[63] Conviction stringency was constructed by residualizing an indicator for conviction (without incarceration) against county-by-year, county-by-month-of-year, and day-of-week fixed effects, then constructing leave-one-out averages at the judge-by-three-year level. Incarceration stringency is similarly constructed over all cases.

incarceration stringency, with the sixth column controlling for dismissal stringency. Again, the loading on incarceration stringency is large and statistically significant, with partial F-statistics between 386 and 468.

Random assignment. As discussed in Section 2.1, within our sample, cases are quasi-randomly assigned to judges within court. There is either actual randomization, or case assignment is done based on scheduling or judge availability.[64] In addition, we confirm empirically that judge stringency does not appear to be meaningfully correlated with case characteristics. In Table 3, in the main paper, we show that case characteristics are strongly correlated with the likelihood of being convicted as well the likelihood of being incarcerated (columns 1 and 3). We then show that case characteristics are largely not correlated with judge conviction stringency (column 2) or incarceration stringency (column 4). For the few instances where covariates that have statistically significant loadings, the predicted difference in stringency tends to be very small (0.02 to 0.04 standard deviations of the residualized stringency instrument).[65] As additional evidence of exogeneity, adding controls for observable defendant characteristics barely changes the first-stage estimates, as seen by comparing columns 2 and 3 and columns 5 and 6 of Table 2 in the main paper.

Exclusion. Our identification strategy relies on the assumption that the conviction stringency instrument affects recidivism outcomes only through its effects on conviction once we control for judges' incarceration stringency, and vice versa. Here we argue that the risk of potential exclusion violations is low. We consider sentence length to be the most important potential violation. For example, if a high-conviction judge also tends to give longer sentences (holding incarceration probability fixed) it would violate exclusion. We test for this by regressing sentence length on our measure of conviction stringency, controlling for incarceration stringency. As shown in Appendix Table D.1, we find no evidence of a violation of the exclusion restriction for conviction. In addition, when we re-estimate the main IV regressions with an additional control for sentence length stringency, we find that the main conclusions are unchanged (see Appendix Figures E.2-E.7).[66]

A judge may influence other aspects of the case, such as probation and parole terms, or fines and fees. While we don't rule these channels out, we don't expect them to be as important. There are a number of large-scale RCTs that have shown probation and parole conditions do not affect recidivism (for a recent review, see Doleac, 2023). There is also a small but growing literature showing that court fines and fees do not affect recidivism (Pager et al., 2022; Finlay et al., 2022; Lieberman et al., 2023). The findings in this literature add confidence that even if judge stringency in conviction and incarceration were correlated with these other factors, they would not bias the results.

[64] In Appendix E, we show that IV estimates are similar when we remove courts where assignment is by judge availability.

[65] We note that conviction stringency appears to be correlated with assault cases. In Appendix E, we show that our results are similar if we drop assault cases, and that people charged with violent crimes are under-represented among compliers.

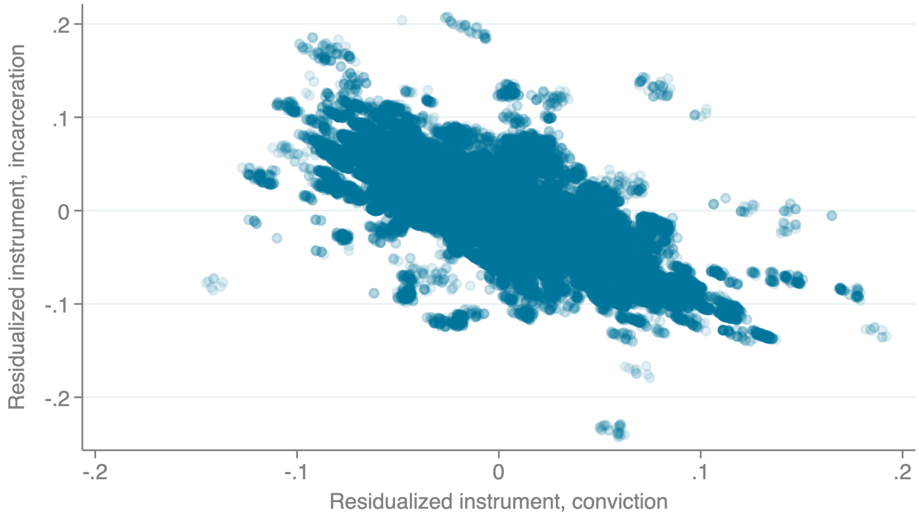
[66] We define sentence length stringency as the tri-yearly leave-one-out average sentence for the judge handling the case, setting sentences to 0 if a person has no carceral sentence and to the sentence length in months if a person is sentenced to a carceral sentence.

We don't expect decisions made at the beginning of the case, such as bail or pre-trial detention, to lead to an exclusion violation. These decisions are made by bail magistrates that have no later influence over the case. Furthermore, it often takes a month or more between the date of arrest and when the defendant arrives in Circuit Court and is assigned a judge. The Circuit Court judge has no influence over these early aspects of the defendant's criminal justice experience.

Although we are comfortable arguing that conviction and incarceration are likely the most important channels by which criminal justice involvement can affect recidivism, we see expanding beyond a trinary model to include these alternatives as an important area of future research. Given the tradeoffs, we have chosen tractability over complexity.

Monotonicity. As discussed previously, one consequence of CPM (and the stronger condition, UPM) is that there will only be one-way flows across any margin (e.g. there can't be some defendants shifting from $T = d$ to $T = c$ while others are shifting from $T = c$ to $T = d$). This is sometimes referred to as 'no defiers' in the binary treatment framework. Here we present some empirical evidence in support of this assumption. Following common practice for binary treatments (see for example Bhuller et al., 2020 or Norris et al., 2021), we conduct split-sample regressions where the data is bifurcated using observed characteristics such as race and gender. Judge stringency is then estimated on each subsample, and the first stage regression is then run on its complement, controlling for stringency along the other margin. If the 'no defiers' condition holds, we would expect positive coefficients for each sub-sample. Appendix Tables D.2 and D.3 report the coefficient on the instrument from split-sample first stage regressions. For both conviction and incarceration, we find positive coefficients on the instrument for all split-sample estimates. Also see Section 6 of the paper where we present a test of the UPM assumption.

Figure D.1: Scatter plot of residualized instruments



Note: This figure presents a scatter plot of the incarceration and conviction instruments. Each instrument is residualized against court and time fixed effects.

Table D.1: 2SLS regressions of sentence length on conviction stringency

	(1) Sent length	(2) Any incar	(3) 6mo	(4) 1y	(5) 2y	(6) 3y	(7) 4y	(8) 5y	(9) 6y	(10) 7y
Pr. convict	-3.58	-0.047	0.063	-0.035	-0.027	-0.025	0.0027	-0.013	-0.0085	-0.00058
	(54.2)	(0.042)	(0.040)	(0.035)	(0.028)	(0.024)	(0.021)	(0.017)	(0.014)	(0.014)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var.	320.446	0.555	0.378	0.204	0.112	0.077	0.060	0.041	0.034	0.029
N	277065	277065	277065	277065	277065	277065	277065	277065	277065	277065

Note: This table shows a regression of various sentence length variables on z_c. The first column uses sentence length as the outcome, the second any incarceration, third to tenth any incarceration greater than 6 months, 1 year, 2 years, 3 years, 4 years, 5 years, 6 years, and 7 years respectively. All regressions control for z_i, race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table D.2: Split sample monotonicity test: conviction

	Zero	One
Any drug charges	0.551	0.205
Any property charges	0.450	0.194
Any violent charges	0.273	0.099
Black	0.296	0.383
Female	0.897	0.165
Prior Conviction	0.295	0.159

Note: This table shows first-stage estimates for the conviction (without incarceration) instrument where, for each regression, the stringency measure is calculated on a specific subpopulation, and the regression is then run on its complement. For example, the "Zero" column of the "Any drug charges" row calculates judge stringency on those without drug charges, then estimates the first stage on those *with* drug charges, and reports the coefficient on the instrument. Regression includes court-by-year fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. The regression also controls for the leave-one-out propensity of the judge to have cases that end in incarceration.

Table D.3: Split sample monotonicity test: incarceration

	Zero	One
Any drug charges	0.545	0.377
Any property charges	0.664	0.316
Any violent charges	0.316	0.232
Black	0.395	0.590
Female	0.719	0.283
Prior Conviction	0.855	0.340

Note: This table shows first-stage estimates for the incarceration instrument where, for each regression, the stringency measure is calculated on a specific subpopulation, and the regression is then run on its complement. For example, the "Zero" column of the "Any drug charges" row calculates judge stringency on those without drug charges, then estimates the first stage on those *with* drug charges, and reports the coefficient on the instrument. Regression includes court-by-year fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Regression also controls for the leave-one-out propensity of the judge to dismiss cases.

E Additional figures and tables: IV analyses

In this appendix, we present a series of additional analyses and robustness tests for our main IV analyses.

E.1 Overview of analyses

Disposition type by offense. Figure E.1 shows the breakdown of disposition types for four common offenses: drugs, fraud, larceny, and assault. These offense categories differ in seriousness and, while the exact breakdown varies, all disposition types are present in each offense type considered.

Residualized instruments. Figure D.1 plots residualized conviction and incarceration instruments against each other. The two instruments are negatively correlated.

Given that the probabilities of all three case outcomes add up to one, this is as expected: an increase in conviction without incarceration can only come from a decrease in either dismissal or incarceration. Importantly, there is substantial variation in z_c across most of the support of z_i and vice versa.

Robustness to sample choice and specification. In Appendix Figures E.2-E.7, we examine how our main estimates for conviction and incarceration change when we alter our sample or specifications, for our 1 year, 1-4 year, and 1-7 year estimates. We consider the following variations:[67]

- Varying which courts are included (keeping courts where judges described a non-random assignment process, or dropping courts where cases are assigned to judges based on availability);
- Changing the required number of cases seen by a judge in our 3-year window (50 or 150 instead of 100);
- Limiting our sample to years 2000-2012, which is the sample in which we can measure recidivism within 7 years;
- Changing what offenses are included: dropping drug cases and offenses types that are not balanced across judges (see Table 3);
- Keeping only offenses that are included in the RD sample;
- Including controls for sentence length stringency or no controls at all.

Generally, our estimates are very close to our main specification (colored in green and denoted by the red dotted line). Although we occasionally lose statistical significance, estimates from the majority of the specifications remain significantly different from zero at the 95% level when our main estimate is also significant. Our main estimates also tend to fall towards the middle of the range of point estimates.

Differential mobility. Our results could be confounded if conviction or incarceration influence the likelihood of moving outside of Virginia, and therefore the likelihood that we would capture their recidivism in our data. Due to data limitations, we cannot test for this in the IV setting. However, for our RD analyses, we can test to see if there is any discontinuity in the likelihood of living in Virginia for those right above/below the cutoff in the incarceration length score and the probation/jail score. We build an indicator for Virginia residency that is equal to one if the person is marked as being in the state of VA in year 5 post-sentencing and year 7 post-sentencing. Missing observations are excluded.[68] As we can see in Appendix Figure E.8, there is no discontinuity at our cutoff score. Notably, in the incarceration-length sample, the share of people remaining in Virginia 5-7 years after the sentencing date ranges from 79-83% at every score.

Conviction propensity and sentence length. In Appendix Table D.1, we regress sentence length on judges' conviction propensity. This table shows that there is no correlation between conviction propensity and sentence length.

[67]The sample and specification changes are detailed in the footnotes of the figures.

[68]If we instead include missings as 0s the results are very similar. Around 7.7% of the sample is missing this information.

Split-sample monotonicity. We next test for monotonicity in judge preferences, following common practice for binary treatments (e.g. Bhuller et al., 2020 or Norris et al., 2021). We conduct split-sample regressions where the data is bifurcated using observed characteristics. Judge stringency is estimated on each subsample, and the first stage regression is then run on its complement. We control for stringency along the other margin. If the 'no defiers' condition holds, we would expect positive coefficients for each sub-sample. Appendix Tables D.2 and D.3 report the coefficient on the instrument from split-sample first-stage regressions. Each row presents a particular case characteristic. For example, the first row breaks our sample into whether a person has a drug charge or does not. The "Zero" column for that row calculates the stringency on the individuals without a drug charge and then estimates the first stage on those with a drug charge, reporting the coefficient on that instrument. The "One" column does the converse of that – calculates the stringency on the individuals with a drug charge and then estimates the first stage on those without a drug charge, reporting the coefficient on that instrument. For both conviction and incarceration, we find positive coefficients on the instrument for all split-sample estimates.

Reduced-form estimates Appendix Table E.1 presents reduced-form estimates, showing the relationship between our outcome variables and the conviction instrument controlling for race, gender, prior conviction, offense type dummies, and year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects as well as the leave-one-out judge incarceration stringency. We find that the instrument positively and significantly affects the year 1, some of the year 1-4, and some of the year 1-7 outcomes. Appendix Table E.2 shows comparable reduced-form estimates for the incarceration results.

Destabilization or "ratcheting up"? We take two strategies to provide suggestive evidence on whether the recidivism effects come from the destabilization channel or the "ratcheting up" effect. First, we look at differences across different stages of the criminal justice process. If each discretionary decision is influenced by the criminal record, then the influence of the record will accumulate as someone advances through the criminal proceedings. If the ratcheting up effect is operative, it may have a larger effect on the more downstream measures of future criminal justice contact, like incarceration, than on the more upstream measures, like new charges. Consistent with this mechanism, we note that in all of our estimates presented in Table 4, the percent changes are larger for more downstream measures of future criminal justice contact.[69]

Second, we consider recidivism across crime types. Following Deshpande and Mueller-Smith (2022), we break out new crimes into income generating crimes or other crimes.[70] If our results are driven by increases in income-generating crime, this would be more consistent with the destabilization channel. Appendix Table E.3 shows that our point estimates are similar for both crime types. The impacts are larger in percent change terms for more downstream measures of future criminal justice contact. Results are simliar if we break out drug crimes from non-drug crimes (Appendix Table E.4).

[69]The fact that conviction increases the probability of future incarceration also indicates that there are direct future financial costs within the criminal justice of these marginal convictions.

[70]Income generating crimes are cases with at least one burglary, drug (excluding drug possession), fraud, larceny, robbery, or prostitution charge.

These analyses are far from definitive, but they provide some suggestive evidence in favor of the "ratcheting up" channel.

2SLS estimates for other subgroups. In Appendix Tables E.5 - E.7, we present 2SLS estimates conditional on various offense categories and sociodemographic characteristics. Appendix Table E.5 separately considers people with or without prior convictions in the last 5 years. We find large effects of conviction for those with no prior felony conviction. Our sample of those with a prior felony conviction is quite small and standard errors are too large to inform us about differences in effect sizes across groups. For incarceration, we find that both groups have similar patterns: short-term incapacitation effects, but no long-term effects, for either group.

We find no substantial differences between Black and White defendants (Appendix Table E.6). We do find some evidence that impacts are larger for people living in ZIP codes with above median poverty rates (Appendix Table E.7). This could be because felony convictions have more consequences in terms of access to relevant social services or housing, or in terms of future criminal justice scrutiny, for poorer people.

Characterizing compliers. Appendix Tables E.8 and E.9 compare compliers for the conviction and incarceration margins to the full sample. The distribution of offenses is mostly similar for compliers to both instruments. They do however differ in some notable ways. Compliers to the conviction instrument are more likely to be female (31% vs 23%) and are more likely to have a property crime charge (45% vs 37%). Moreover, they are less likely to have a prior conviction (12% vs 18%), less likely to have a violent charge than the general sample (9% vs 20%), and less likely to have charges that fall into the other category (5% vs 15%). Compliers to the incarceration instrument are slightly more similar to the full sample, but exhibit some of the same notable differences. First, prior conviction rates and share of women are more similar, 21% vs 18% and 23% vs 23%, respectively. For property charges and violent charges we continue to see disparities with 45% vs 37% having a property charge and 10% vs 20% having a violent charge.

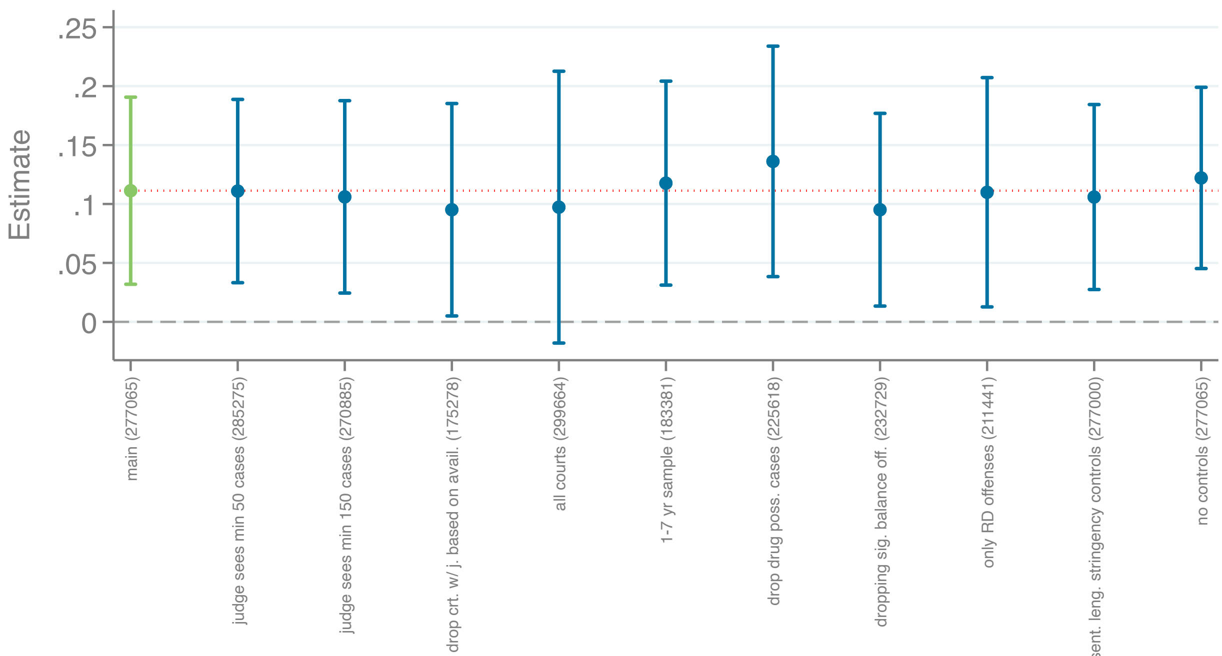
Testing between models of judge decision-making. Appendix Table E.10 shows the first and second tests described in Subsection 6.1 for predicted recidivism. In Panel A of Appendix Table E.10, we regress predicted recidivism on z_c while controlling for z_i and court and time fixed effects and limiting the sample to those convicted of incarceration. The test is against the null that the coefficient on z_c is zero, since conviction stringency should not affect the characteristics of those incarcerated. In Panel B, using the dismissed sample, we regress predicted recidivism on z_i while controlling for z_d and court and time fixed effects. Similar to the first test, we test against the null that the coefficient on z_i is zero. Using predicted recidivism as an index of characteristics, we reject the null in both tests.

E.2 Appendix Figures: 2SLS analyses

Figure E.1: Dismissed, convicted, and incarcerated percentages by offenses

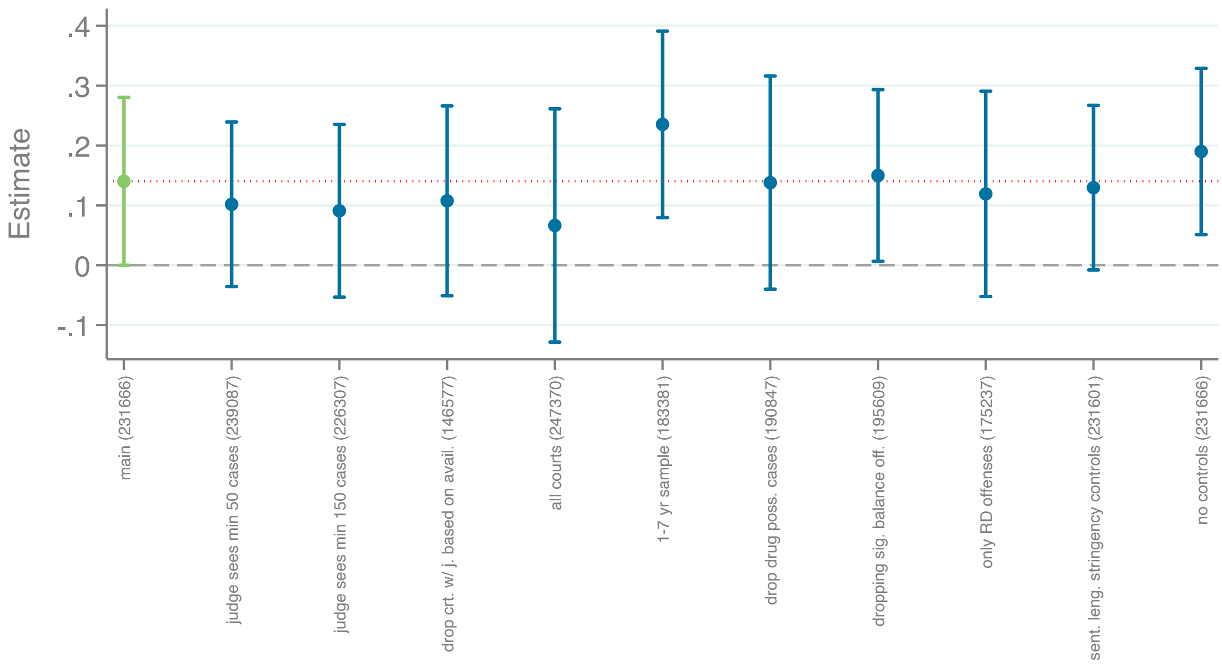
Note: This figure shows the variation in dismissal, conviction, and incarceration by four offense categories. The top left depicts fraud cases, the top right larceny, the bottom left assault, and the bottom right drugs. There is variation in the percent of cases dismissed, convicted, and incarcerated within each offense.

Figure E.2: Robustness for 2SLS conviction results: recidivism in year 1



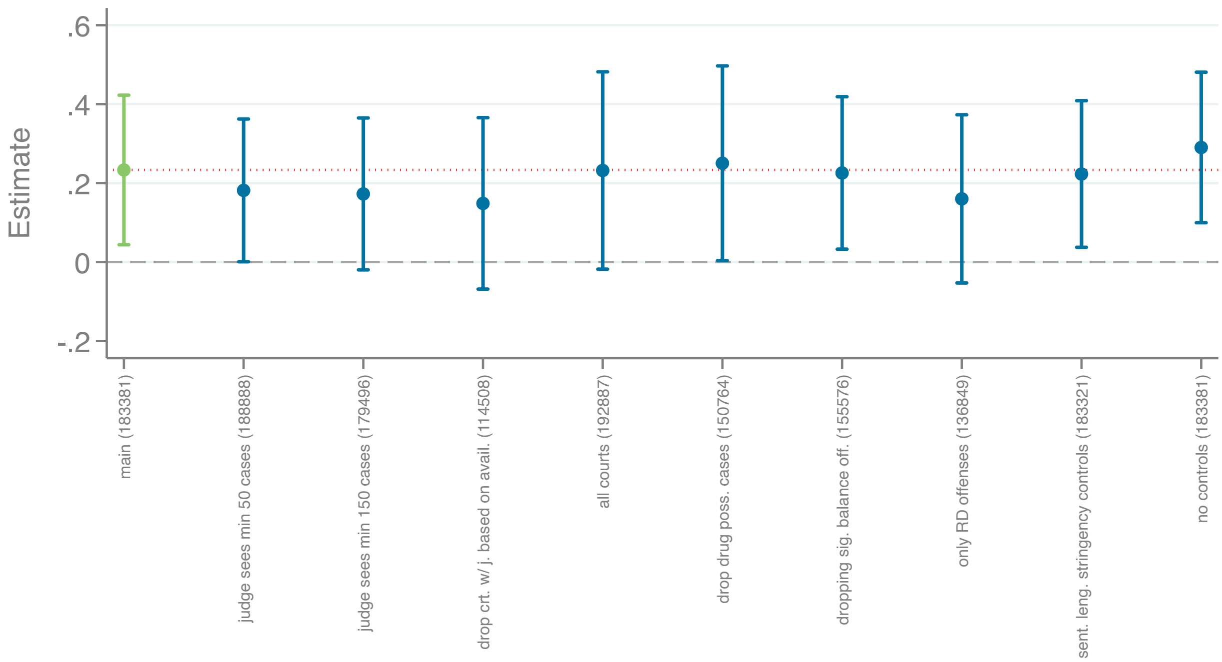
Note: This figure shows various estimates of the impact of conviction on recidivism within the first year after sentencing. Sample size is noted in parentheses and the main estimate is highlighted in green; 95% confidence intervals are shown. The red dotted line is located at the height of the main estimate and the dashed gray line is located at 0. The sample restrictions on the estimates are the following: (1) Our main specification. (2) Requiring that a judge sees a minimum of 50 cases in our 3 year window. (3) Requiring that a judge sees a minimum of 150 cases in our 3 year window. (4) Dropping courts that use judges based on availability. (5) Keeping courts where clerks described an assignment process that seemed non-random. (6) Using the 1-7 year sample restriction on year 1 and years 1-4. (7) Dropping any cases that relate to drug possession. (8) Dropping any offenses that are significant in our balance tests. (9) Restricting to offenses that are observed in the RD sample. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls.

Figure E.3: Robustness for 2SLS conviction results: recidivism in years 1-4



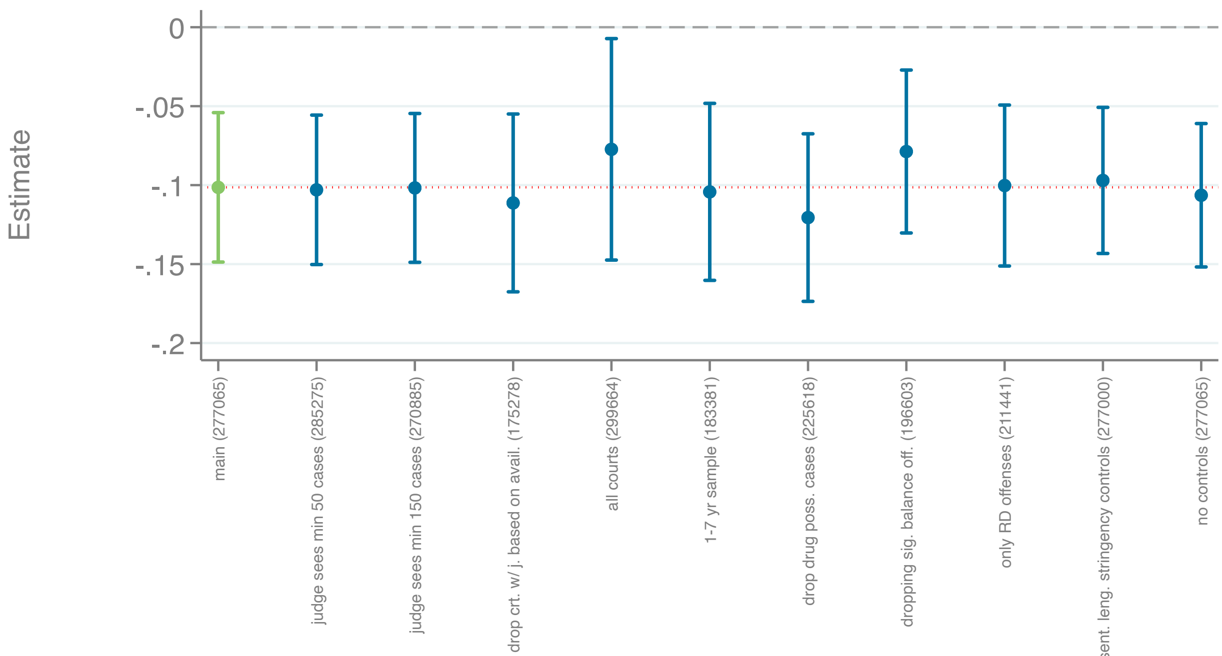
Note: This figure shows various estimates of the impact of conviction on recidivism within years 1-4 after sentencing. Sample size is noted in parentheses and the main estimate is highlighted in green; 95% confidence intervals are shown. The red dotted line is located at the height of the main estimate and the dashed gray line is located at 0. The sample restrictions on the estimates are the following: (1) Our main specification. (2) Requiring that a judge sees a minimum of 50 cases in our 3 year window. (3) Requiring that a judge sees a minimum of 150 cases in our 3 year window. (4) Dropping courts that use judges based on availability. (5) Keeping courts where clerks described an assignment process that seemed non-random. (6) Using the 1-7 year sample restriction on year 1 and years 1-4. (7) Dropping any cases that relate to drug possession. (8) Dropping any offenses that are significant in our balance tests. (9) Restricting to offenses that are observed in the RD sample. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls.

Figure E.4: Robustness for 2SLS conviction results: recidivism in years 1-7



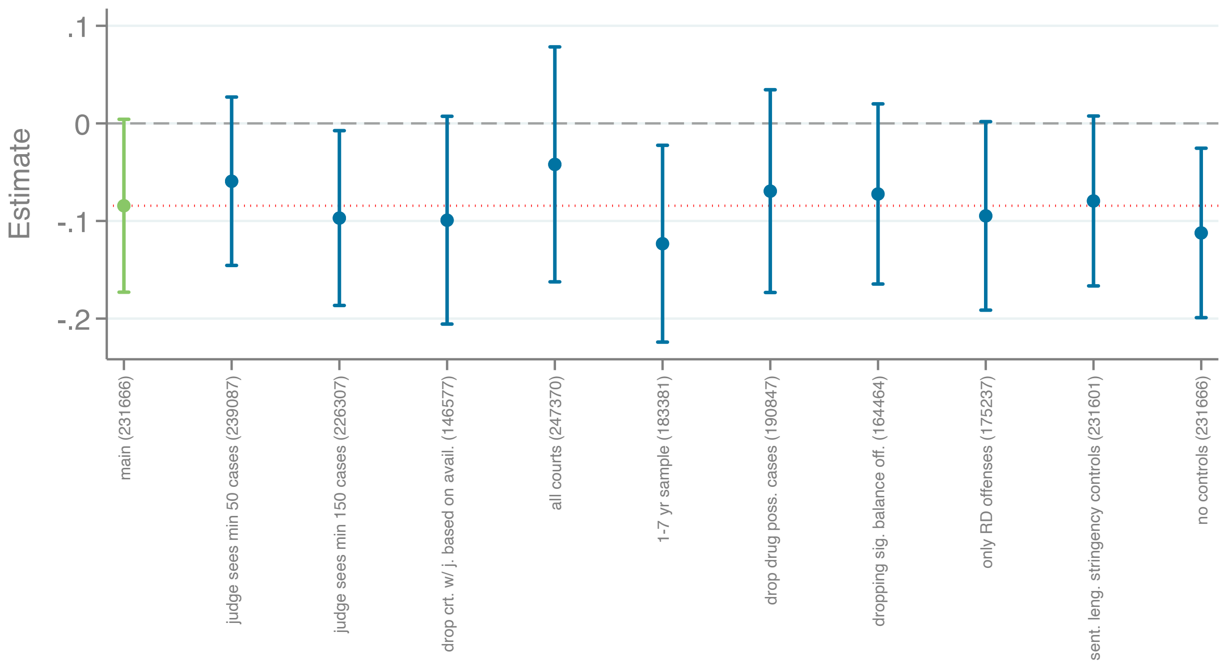
Note: This figure shows various estimates of the impact of conviction on recidivism within years 1-7 after sentencing. Sample size is noted in parentheses and the main estimate is highlighted in green; 95% confidence intervals are shown. The red dotted line is located at the height of the main estimate and the dashed gray line is located at 0. The sample restrictions on the estimates are the following: (1) Our main specification. (2) Requiring that a judge sees a minimum of 50 cases in our 3 year window. (3) Requiring that a judge sees a minimum of 150 cases in our 3 year window. (4) Dropping courts that use judges based on availability. (5) Keeping courts where clerks described an assignment process that seemed non-random. (6) Dropping any cases that relate to drug possession. (7) Dropping any offenses that are significant in our balance tests. (8) Restricting to offenses that are observed in the RD sample. (9) Including a sentence length stringency instrument control. (10) Main specification without any of our controls.

Figure E.5: Robustness for 2SLS incarceration results: recidivism in year 1



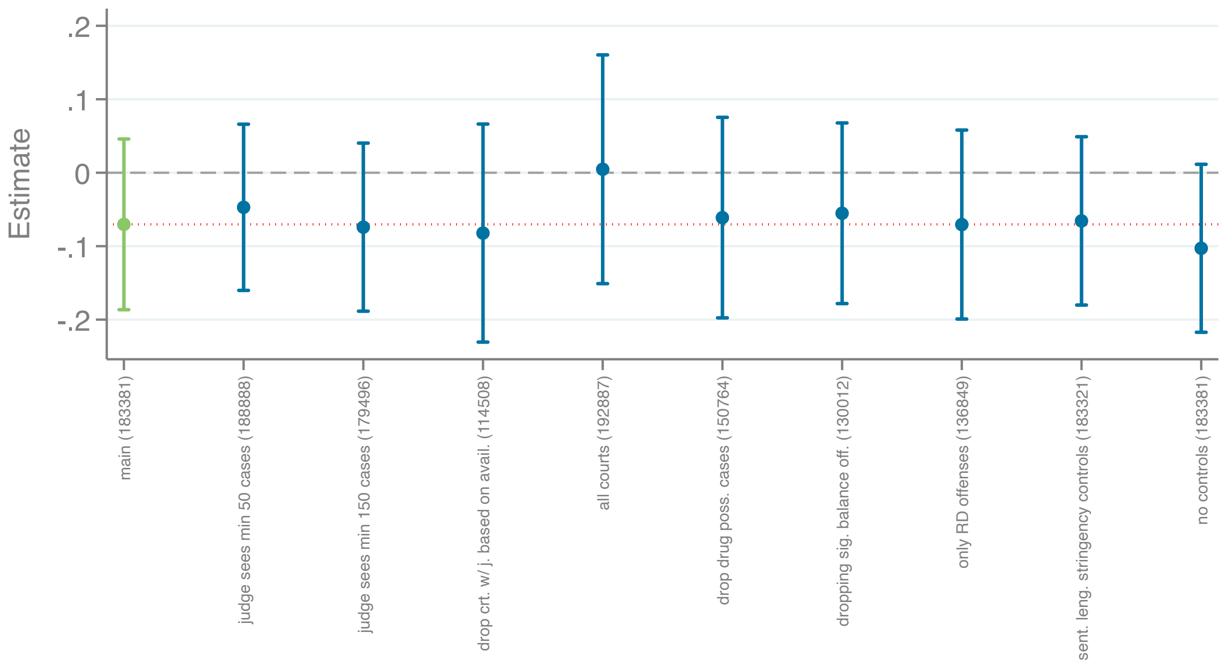
Note: This figure shows various estimates of the impact of incarceration on recidivism within the first year after sentencing. Sample size is noted in parentheses and the main estimate is highlighted in green; 95% confidence intervals are shown. The red dotted line is located at the height of the main estimate and the dashed gray line is located at 0. The sample restrictions on the estimates are the following: (1) Our main specification. (2) Requiring that a judge sees a minimum of 50 cases in our 3 year window. (3) Requiring that a judge sees a minimum of 150 cases in our 3 year window. (4) Dropping courts that use judges based on availability. (5) Keeping courts where clerks described an assignment process that seemed non-random. (6) Using the 1-7 year sample restriction on year 1 and years 1-4. (7) Dropping any cases that relate to drug possession. (8) Dropping any offenses that are significant in our balance tests. (9) Restricting to offenses that are observed in the RD sample. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls.

Figure E.6: Robustness for 2SLS incarceration results: recidivism in years 1-4



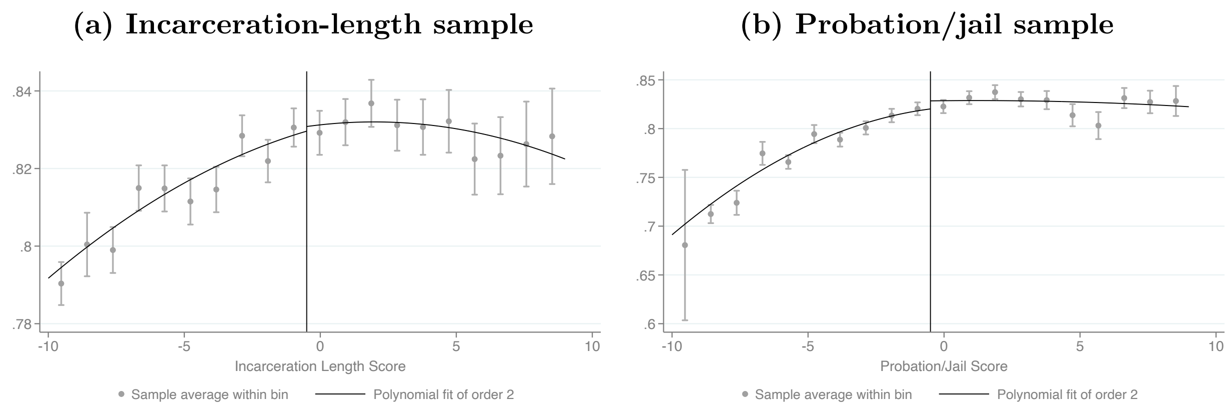
Note: This figure shows various estimates of the impact of incarceration on recidivism within years 1-4 after sentencing. Sample size is noted in parentheses and the main estimate is highlighted in green; 95% confidence intervals are shown. The red dotted line is located at the height of the main estimate and the dashed gray line is located at 0. The sample restrictions on the estimates are the following: (1) Our main specification. (2) Requiring that a judge sees a minimum of 50 cases in our 3 year window. (3) Requiring that a judge sees a minimum of 150 cases in our 3 year window. (4) Dropping courts that use judges based on availability. (5) Keeping courts where clerks described an assignment process that seemed non-random. (6) Using the 1-7 year sample restriction on year 1 and years 1-4. (7) Dropping any cases that relate to drug possession. (8) Dropping any offenses that are significant in our balance tests. (9) Restricting to offenses that are observed in the RD sample. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls.

Figure E.7: Robustness for 2SLS incarceration results: recidivism in years 1-7



Note: This figure shows various estimates of the impact of incarceration on recidivism within years 1-7 after sentencing. Sample size is noted in parentheses and the main estimate is highlighted in green; 95% confidence intervals are shown. The red dotted line is located at the height of the main estimate and the dashed gray line is located at 0.The sample restrictions on the estimates are the following: (1) Our main specification. (2) Requiring that a judge sees a minimum of 50 cases in our 3 year window. (3) Requiring that a judge sees a minimum of 150 cases in our 3 year window. (4) Dropping courts that use judges based on availability. (5) Keeping courts where clerks described an assignment process that seemed non-random. (6) Dropping any cases that relate to drug possession. (7) Dropping any offenses that are significant in our balance tests. (8) Restricting to offenses that are observed in the RD sample. (9) Including a sentence length stringency instrument control. (10) Main specification without any of our controls.

Figure E.8: Testing for discontinuities in Virginia residency



Note: The outcome variable here is a flag indicating that the person is still residing in Virginia 5-7 years after their sentencing date, based on data obtained from a private vendor. Panel (a) is restricted to the RD incarceration-length sample; panel (b) is restricted to the RD probation/jail sample. There is no discontinuity across either threshold. People whose residency information is missing (7.7% of the sample) were excluded from the analysis.

E.3 Appendix Tables: 2SLS analyses

Table E.1: Reduced form conviction estimates

	Year 1	Year 2-4	Year 5-4	Year 1-4	Year 1-7
	RF	RF	RF	RF	RF
Fut. charge	0.063***	0.028	0.045	0.082**	0.138**
	(0.023)	(0.038)	(0.044)	(0.041)	(0.055)
Fut. conviction	0.077***	0.041	0.032	0.119***	0.176***
	(0.021)	(0.037)	(0.042)	(0.040)	(0.053)
Fut. incarceration	0.062***	0.011	-0.015	0.072**	0.126***
	(0.019)	(0.032)	(0.034)	(0.036)	(0.048)
Observations	277065	231666	183381	231666	183381

Note: This table shows estimates from reduced form regressions of recidivism on z_c. The five columns report results for five recidivism time ranges (1 year, 2-4 years, 5-7 years, 1-4 years, and 1-7 years). The samples are restricted to cases in which the full recidivism time window is observed. All regressions control for z_i, race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. The table reports the estimated impact of conviction. The first row is for any future felony charge, the second row is for any future conviction, and the third row is for any future incarceration. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table E.2: Reduced form incarceration estimates

	Year 1	Year 2-4	Year 5-4	Year 1-4	Year 1-7
	RF	RF	RF	RF	RF
Fut. charge	-0.062***	-0.006	0.002	-0.052*	-0.042
	(0.015)	(0.025)	(0.024)	(0.028)	(0.036)
Fut. conviction	-0.068***	-0.018	0.013	-0.075***	-0.064*
	(0.014)	(0.024)	(0.023)	(0.027)	(0.035)
Fut. incarceration	-0.050***	0.007	0.032	-0.033	-0.018
	(0.012)	(0.021)	(0.019)	(0.024)	(0.031)
Observations	277065	231666	183381	231666	183381

Note: This table shows estimates from reduced form regressions of recidivism on z_i. The five columns report results for five recidivism time ranges (1 year, 2-4 years, 5-7 years, 1-4 years, and 1-7 years). The samples are restricted to cases in which the full recidivism time window is observed. All regressions control for z_d, race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. The table reports the estimated impact of incarceration. The first row is for any future felony charge, the second row is for any future conviction, and the third row is for any future incarceration. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table E.3: Income-generating vs non-income-generating recidivism

	Income generating			Non-income generating		
	Year 1	Year 1-4	Year 1-7	Year 1	Year 1-4	Year 1-7
Panel A: Conviction						
Fut. charge	0.058**	0.099*	0.131*	0.064**	0.115*	0.128
	(0.028)	(0.059)	(0.075)	(0.030)	(0.063)	(0.083)
Fut. conviction	0.061**	0.159***	0.185**	0.071***	0.133**	0.117
	(0.026)	(0.057)	(0.074)	(0.028)	(0.060)	(0.081)
Fut. incarceration	0.046**	0.083*	0.084	0.058**	0.111**	0.150**
	(0.022)	(0.048)	(0.065)	(0.023)	(0.051)	(0.067)
Ctrl Mean: fut. charge	0.040	0.148	0.219	0.041	0.151	0.226
Ctrl Mean: fut. conv.	0.036	0.134	0.200	0.035	0.133	0.202
Ctrl Mean: fut. incar.	0.028	0.106	0.160	0.026	0.101	0.153
Panel B: Incarceration						
Fut. charge	-0.046***	-0.026	-0.057	-0.043***	-0.066*	-0.026
	(0.015)	(0.036)	(0.047)	(0.016)	(0.039)	(0.052)
Fut. conviction	-0.050***	-0.039	-0.069	-0.047***	-0.070*	-0.006
	(0.015)	(0.036)	(0.046)	(0.016)	(0.037)	(0.050)
Fut. incarceration	-0.041***	-0.015	-0.021	-0.034***	-0.029	0.012
	(0.012)	(0.031)	(0.042)	(0.013)	(0.031)	(0.040)
Ctrl Mean: fut. charge	0.050	0.148	0.203	0.047	0.142	0.199
Ctrl Mean: fut. conv.	0.045	0.134	0.185	0.040	0.124	0.176
Ctrl Mean: fut. incar.	0.032	0.099	0.140	0.027	0.087	0.125
Observations	285189	235021	184726	285189	235021	184726

Note: The first panel of this table shows 2SLS estimates of the impact of conviction vs dismissal on future recidivism. In the first three columns, recidivism is defined in reference to new income-generating felony charges, in the last three columns recidivism is defined in reference to new non-income generating charges. The second panel is similar except it shows 2SLS estimates of the impact of incarceration vs conviction. The columns report results for three recidivism time ranges (1 year, 1-4 years, and 1-7 years). Each time period restricts the sample to cases for which the full recidivism time period is observed. All regressions control for stringency on the other margin (i.e. z_i for the conviction specification), race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01.

Table E.4: Drug vs non-drug recidivism

	Drug Charges			Non-Drug Charges		
	Year 1	Year 1-4	Year 1-7	Year 1	Year 1-4	Year 1-7
Panel A: Conviction						
Fut. charge	0.162**	0.183	0.160	0.076	0.123	0.298**
	(0.065)	(0.115)	(0.140)	(0.050)	(0.090)	(0.121)
Fut. conviction	0.138**	0.204*	0.204	0.129***	0.208**	0.369***
	(0.057)	(0.106)	(0.134)	(0.047)	(0.089)	(0.118)
Fut. incarceration	0.114**	0.101	0.130	0.103**	0.138*	0.272***
	(0.049)	(0.098)	(0.128)	(0.042)	(0.077)	(0.103)
Ctrl Mean: fut. charge	0.072	0.251	0.360	0.067	0.233	0.331
Ctrl Mean: fut. conv.	0.064	0.226	0.331	0.059	0.213	0.306
Ctrl Mean: fut. incar.	0.048	0.177	0.260	0.046	0.169	0.246
Panel B: Incarceration						
Fut. charge	-0.060	-0.094	-0.046	-0.120***	-0.085	-0.090
	(0.048)	(0.086)	(0.109)	(0.028)	(0.054)	(0.069)
Fut. conviction	-0.072*	-0.126	-0.082	-0.129***	-0.125**	-0.123*
	(0.043)	(0.082)	(0.109)	(0.026)	(0.053)	(0.066)
Fut. incarceration	-0.062*	-0.013	-0.017	-0.093***	-0.071	-0.042
	(0.037)	(0.077)	(0.102)	(0.023)	(0.046)	(0.058)
Ctrl Mean: fut. charge	0.087	0.239	0.317	0.083	0.225	0.296
Ctrl Mean: fut. conv.	0.076	0.215	0.290	0.072	0.203	0.271
Ctrl Mean: fut. incar.	0.053	0.157	0.216	0.051	0.150	0.206
Observations	88553	72258	57248	188512	159408	126133

Note: The first panel of this table shows 2SLS estimates of the impact of conviction vs dismissal on future recidivism. In the first three columns, recidivism is defined in reference to new drug charges, in the last three columns recidivism is defined in reference to new non-drug charges. The second panel is similar except it shows 2SLS estimates of the impact of incarceration. The columns report results for three recidivism time ranges (1 year, 1-4 years, and 1-7 years). Each time period restricts the sample to cases for which the full recidivism window is observed. All regressions control for stringency on the opposite margin (i.e. z_i in the conviction specification) race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table E.5: 2SLS estimates for those with/without prior felony convictions

	Priors			No Priors		
	Year 1	Year 1-4	Year 1-7	Year 1	Year 1-4	Year 1-7
Panel A: Conviction						
Fut. charge	0.187	0.334	0.316	0.096**	0.128*	0.228**
	(0.169)	(0.358)	(0.444)	(0.041)	(0.072)	(0.093)
Fut. conviction	0.233	0.329	0.352	0.118***	0.198***	0.293***
	(0.156)	(0.343)	(0.448)	(0.038)	(0.070)	(0.090)
Fut. incarceration	0.230*	0.354	0.443	0.090***	0.106*	0.190**
	(0.139)	(0.311)	(0.431)	(0.032)	(0.060)	(0.078)
Ctrl Mean: fut. charge	0.078	0.318	0.473	0.066	0.218	0.306
Ctrl Mean: fut. conv.	0.070	0.293	0.444	0.058	0.197	0.280
Ctrl Mean: fut. incar.	0.055	0.236	0.362	0.044	0.155	0.222
Panel B: Incarceration						
Fut. charge	-0.174***	-0.099	-0.021	-0.079***	-0.080	-0.073
	(0.057)	(0.100)	(0.131)	(0.027)	(0.050)	(0.062)
Fut. conviction	-0.183***	-0.177*	-0.106	-0.089***	-0.109**	-0.097
	(0.054)	(0.099)	(0.127)	(0.025)	(0.049)	(0.060)
Fut. incarceration	-0.145***	-0.111	-0.021	-0.063***	-0.041	-0.023
	(0.048)	(0.090)	(0.120)	(0.021)	(0.042)	(0.053)
Ctrl Mean: fut. charge	0.123	0.373	0.499	0.079	0.211	0.278
Ctrl Mean: fut. conv.	0.111	0.345	0.471	0.069	0.189	0.253
Ctrl Mean: fut. incar.	0.082	0.272	0.379	0.048	0.137	0.188
Observations	49926	40738	31504	227139	190928	151877

Note: The first panel of this table shows 2SLS estimates of the impact of conviction on future recidivism for those with/without a prior felony conviction within 5 years. The second panel is similar except it shows 2SLS estimates of the impact of incarceration. The columns report results for three recidivism time ranges (1 year, 1-4 years, and 1-7 years). Each time period restricts the sample to cases for which the full recidivism window is observed. All regressions control for stringency on the opposite margin (i.e. z_i in the conviction specification) race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E.6: 2SLS estimates for Black and non-Black defendants

	Black			Non-Black		
	Year 1	Year 1-4	Year 1-7	Year 1	Year 1-4	Year 1-7
Panel A: Conviction						
Fut. charge	0.089	0.152	0.238	0.135**	0.149	0.242**
	(0.058)	(0.106)	(0.147)	(0.057)	(0.095)	(0.114)
Fut. conviction	0.117**	0.216**	0.335**	0.156***	0.205**	0.274**
	(0.056)	(0.103)	(0.146)	(0.052)	(0.093)	(0.111)
Fut. incarceration	0.108**	0.128	0.260**	0.105**	0.125	0.182*
	(0.050)	(0.087)	(0.126)	(0.045)	(0.084)	(0.102)
Ctrl Mean: fut. charge	0.072	0.258	0.369	0.063	0.211	0.298
Ctrl Mean: fut. conv.	0.063	0.233	0.338	0.058	0.195	0.277
Ctrl Mean: fut. incar.	0.049	0.185	0.271	0.043	0.152	0.220
Panel B: Incarceration						
Fut. charge	-0.138***	-0.101	-0.114	-0.065**	-0.063	-0.027
	(0.036)	(0.066)	(0.086)	(0.032)	(0.061)	(0.082)
Fut. conviction	-0.133***	-0.131**	-0.140	-0.091***	-0.110*	-0.072
	(0.035)	(0.065)	(0.085)	(0.031)	(0.060)	(0.079)
Fut. incarceration	-0.113***	-0.081	-0.071	-0.050*	-0.022	0.014
	(0.029)	(0.055)	(0.074)	(0.027)	(0.054)	(0.072)
Ctrl Mean: fut. charge	0.091	0.254	0.335	0.076	0.203	0.267
Ctrl Mean: fut. conv.	0.079	0.226	0.305	0.068	0.185	0.246
Ctrl Mean: fut. incar.	0.057	0.170	0.233	0.046	0.133	0.183
Observations	154724	130736	104217	122341	100930	79164

Note: The first panel of this table shows 2SLS estimates of the impact of conviction on future recidivism for Black and non-Black defendants. The second panel is similar except it shows 2SLS estimates of the impact of incarceration. The columns report results for three recidivism time ranges (1 year, 1-4 years, and 1-7 years). Each time period restricts the sample to cases for which the full recidivism window is observed. All regressions control for stringency on the opposite margin (i.e. z_i in the conviction specification) race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table E.7: 2SLS estimates for those from zip codes above and below median poverty level

	Above median poverty zip			Below median poverty zip		
	Year 1	Year 1-4	Year 1-7	Year 1	Year 1-4	Year 1-7
Panel A: Conviction						
Fut. charge	0.137*	0.225*	0.351**	0.051	-0.027	0.055
	(0.078)	(0.124)	(0.171)	(0.048)	(0.093)	(0.123)
Fut. conviction	0.122*	0.280**	0.383**	0.102**	0.033	0.123
	(0.071)	(0.123)	(0.157)	(0.046)	(0.090)	(0.121)
Fut. incarceration	0.068	0.188*	0.296**	0.094**	-0.038	0.042
	(0.057)	(0.110)	(0.148)	(0.043)	(0.079)	(0.102)
Ctrl Mean: fut. charge	0.079	0.274	0.390	0.064	0.224	0.318
Ctrl Mean: fut. conv.	0.069	0.247	0.359	0.057	0.205	0.295
Ctrl Mean: fut. incar.	0.053	0.196	0.288	0.043	0.161	0.235
Panel B: Incarceration						
Fut. charge	-0.116***	-0.063	0.009	-0.078***	-0.005	-0.018
	(0.040)	(0.070)	(0.086)	(0.030)	(0.061)	(0.080)
Fut. conviction	-0.110***	-0.090	-0.001	-0.095***	-0.058	-0.063
	(0.039)	(0.067)	(0.081)	(0.029)	(0.060)	(0.080)
Fut. incarceration	-0.081**	-0.063	0.041	-0.072***	0.021	0.001
	(0.032)	(0.060)	(0.076)	(0.026)	(0.054)	(0.072)
Ctrl Mean: fut. charge	0.097	0.266	0.351	0.080	0.217	0.286
Ctrl Mean: fut. conv.	0.084	0.239	0.321	0.070	0.198	0.264
Ctrl Mean: fut. incar.	0.059	0.178	0.246	0.049	0.144	0.196
Observations	114383	94781	73472	114441	94237	73532

Note: The first panel of this table shows 2SLS estimates of the impact of conviction on future recidivism for those who live in ZIP codes where the percent earning under 25K (percent in poverty) is above/below median. The second panel is similar except it shows 2SLS estimates of the impact of incarceration. The columns report results for three recidivism time ranges (1 year, 1-4 years, and 1-7 years). Each time period restricts the sample to cases for which the full recidivism window is observed. All regressions control for stringency on the opposite margin (i.e. z_i in the conviction specification) race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table E.8: Complier characteristics (conviction)

	Pr(X=x)	Pr(X=x|complier)	$\frac{Pr(X=x\|complier)}{Pr(X=x)}$
Prior conviction	0.180	0.122	0.674
	(0.003)	(0.030)	(0.165)
Female	0.225	0.312	1.387
	(0.003)	(0.036)	(0.156)
Black	0.558	0.581	1.041
	(0.012)	(0.037)	(0.064)
Has misdemeanor	0.098	0.095	0.968
	(0.005)	(0.023)	(0.231)
Drugs	0.320	0.318	0.995
	(0.006)	(0.033)	(0.100)
Property	0.374	0.451	1.207
	(0.006)	(0.040)	(0.103)
Violent	0.201	0.094	0.468
	(0.004)	(0.025)	(0.126)
Other	0.153	0.048	0.315
	(0.002)	(0.024)	(0.155)

Note: This table shows the characteristics of compliers for our 2SLS conviction analysis. The first column reports average characteristics for the full 2SLS sample. The second column reports the estimated average coefficients for compliers. The third column reports the ratio of column 2 to column 1. Standard errors are calculated via bootstrap using 500 bootstrap samples.

Table E.9: Complier characteristics (incarceration)

	Pr(X=x)	Pr(X=x|complier)	$\frac{Pr(X=x\|complier)}{Pr(X=x)}$
Prior conviction	0.180	0.209	1.160
	(0.003)	(0.019)	(0.103)
Female	0.225	0.226	1.005
	(0.003)	(0.022)	(0.097)
Black	0.558	0.531	0.951
	(0.012)	(0.023)	(0.038)
Has misdemeanor	0.098	0.080	0.818
	(0.005)	(0.016)	(0.160)
Drugs	0.320	0.288	0.902
	(0.006)	(0.022)	(0.066)
Property	0.374	0.453	1.212
	(0.006)	(0.026)	(0.066)
Violent	0.201	0.100	0.495
	(0.004)	(0.022)	(0.110)
Other	0.153	0.128	0.840
	(0.002)	(0.017)	(0.110)

Note: This table shows the characteristics of compliers for our 2SLS incarceration analysis. The first column reports average characteristics for the full 2SLS sample. The second column reports the estimated average coefficients for compliers. The third column reports the ratio of column 2 to column 1. Standard errors are calculated via bootstrap using 500 bootstrap samples.

Table E.10: Testing the models with predicted recidivism

	Pred. recid. within 1 year	Pred. recid. 1-4 years	Pred. recid. 1-7 years
Panel A: Ordered			
Pr. convict	0.010***	0.031***	0.038***
	(0.0030)	(0.0079)	(0.0095)
Mean Dep. Var.	0.087	0.230	0.286
N	153692	153692	153692
Panel B: Sequential and ordered			
Pr. incar	-0.0096***	-0.024***	-0.028***
	(0.0032)	(0.0085)	(0.010)
Mean Dep. Var.	0.084	0.212	0.263
N	44114	44114	44114

Note: Predicted recidivism variables are created by taking the fitted values from a regression of recidivism after release on controls for demographics, charge, criminal record, and month, year-by-court, court-by-month-of-year, and day-of-week FE. For Panel A, we restrict to the incarcerated sample and regress predicted recidivism on conviction stringency controlling for incarceration stringency and court-by-time fixed effects. For Panel B, we restrict to the dismissed sample and regress predicted recidivism on incarceration stringency controlling for dismissal stringency and court-by-time fixed effects. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

F Additional derivations and results

F.1 2SLS with two endogenous variables

Here we briefly discuss why our specification, which instruments for a binary treatment indicator (such as T_c) with one stringency (such as Z_c) while controlling for another stringency (such as Z_i), should have the same estimand as running a single 2SLS regression with two endogenous treatment variables and both stringencies. In the main paper, we consider the following specification:

$$T_c = \delta_0 + \delta_1 Z_c + \delta_2 Z_i + U$$
$$Y = \gamma_0 + \gamma_1 T_c + \gamma_2 Z_i + V$$

Note that asymptotically, $\delta_1 = 0$, $\delta_1 = 1$, and $\delta_2 = 0$. Thus, γ_1 should be equal to γ_1' in the following regression:

$$Y = \gamma_0' + \gamma_1' Z_c + \gamma_2' Z_i + V'$$

Consider now a specification in which both endogenous variables, T_c and T_i, are instrumented for in the same second stage regression:

$$T_c = \delta_0 + \delta_1 Z_c + \delta_2 Z_i + U$$
$$T_i = \omega_0 + \omega_1 Z_c + \omega_2 Z_i + U$$
$$Y = \gamma_0'' + \gamma_1'' T_c + \gamma_2'' T_i + V''$$

By similar logic, $\omega_0 = 0$, $\omega_1 = 0$, and $\omega_2 = 1$ asymptotically. Thus, $\gamma_1 = \gamma_1' = \gamma_1''$ and $\gamma_2 = \gamma_2' = \gamma_2''$.

In practice, our first-stage coefficients are not precisely zero or one, as is common in the applied literature, yet both approaches produce similar estimates. Table F.1 shows that, when running 2SLS with two instruments and two endogenous variables, our estimates are similar to those in the main paper and we reach similar conclusions. Note that in these 2SLS and OLS regressions we replace T_c with $T_{\setminus D}$ (i.e. the conviction instrument dummy that remains equal to one for those incarcerated) so that the loading on T_i can be interpreted as the change relative to $T = C$ rather than $T = D$.

Table F.1: Two instruments and two endogenous variables

	Year 1		Year 2-4		Year 5-7		Year 1-4		Year 1-7	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Convict: fut. charge	0.001	0.112***	0.008***	0.050	0.006**	0.082	0.008***	0.142*	0.011***	0.243**
	(0.002)	(0.043)	(0.003)	(0.067)	(0.002)	(0.078)	(0.003)	(0.073)	(0.004)	(0.100)
Incar: fut. charge	-0.021***	-0.104***	0.012***	-0.011	0.025***	0.004	-0.003	-0.087*	0.023***	-0.071
	(0.001)	(0.025)	(0.002)	(0.040)	(0.002)	(0.040)	(0.002)	(0.046)	(0.003)	(0.059)
Convict: fut. conv.	0.003*	0.137***	0.010***	0.072	0.007***	0.059	0.011***	0.207***	0.014***	0.309***
	(0.002)	(0.040)	(0.003)	(0.064)	(0.002)	(0.074)	(0.003)	(0.072)	(0.004)	(0.098)
Incar: fut. conv.	-0.018***	-0.114***	0.013***	-0.030	0.023***	0.021	-0.001	-0.126***	0.022***	-0.106*
	(0.001)	(0.024)	(0.002)	(0.040)	(0.002)	(0.039)	(0.002)	(0.045)	(0.003)	(0.058)
Convict: fut. incar.	0.003**	0.111***	0.009***	0.020	0.005**	-0.023	0.010***	0.126**	0.012***	0.224***
	(0.001)	(0.035)	(0.002)	(0.057)	(0.002)	(0.059)	(0.002)	(0.062)	(0.003)	(0.086)
Incar: fut. incar.	-0.009***	-0.084***	0.016***	0.012	0.021***	0.053	0.008***	-0.056	0.027***	-0.030
	(0.001)	(0.020)	(0.002)	(0.035)	(0.002)	(0.032)	(0.002)	(0.040)	(0.003)	(0.051)
Ctrl Mean: fut. charge	0.083	0.083	0.165	0.165	0.129	0.129	0.223	0.223	0.297	0.297
Ctrl Mean: fut. conv.	0.071	0.071	0.144	0.144	0.114	0.114	0.197	0.197	0.268	0.268
Ctrl Mean: fut. incar.	0.050	0.050	0.105	0.105	0.083	0.083	0.146	0.146	0.204	0.204
Observations	277,065	277,065	231,666	231,666	183,381	183,381	231,666	231,666	183,381	183,381

Note: This table shows regression estimates of the impacts of conviction and incarceration on future recidivism. The five columns report results for five time ranges (1 year, 2-4 years, 5-7 years, 1-4 years, and 1-7 years). For each panel we report ordinary least squares (OLS) and instrumental variable (IV) estimates with two instruments and two endogenous variables. Each time period restricts the sample to cases for which the full time period is observed. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month-of-year fixed effects, and day-of-week fixed effects. The first six rows report the estimated impact of conviction or incarceration on different measures of recidivism. The first two rows are for any future charge, the second two rows are for any future conviction, and the third two rows are for any future incarceration. For the OLS estimates, we regress our measures of recidivism on having a conviction (regardless of incarceration status) controlling for incarceration. The estimates presented are the coefficient on the conviction variable. For the IV estimates, this provides an estimate of the impacts of conviction compared to dismissal for the set of compliers at that margin and incarceration compared to conviction for the set of compliers at the other margin. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

F.2 Binary treatment

Consider an attempt to estimate the impacts of incarceration vs non-incarceration using the following 2SLS specification:

$$T_i = \delta_0 + \delta_1 Z_i + U$$
$$Y = \gamma_0 + \gamma_1 T_i + V$$

This specification is similar to Equations (1) and (2) from the main text, but does not include judge dismissal stringency as a control. Under the standard LATE assumptions, γ_1 will not yield a weighted average of LATEs of incarceration vs non-incarceration, since an increase in Z_i could generate flows between dismissal and conviction in the non-incarcerated group if Z_i and Z_c are correlated, which is likely given that $Z_i = 1 - (Z_c + Z_d)$ by construction.

G Additional details for multinomial model with heterogeneous effects

This appendix discusses the application of Mountjoy (2022) in our setting. First, we describe the identification and estimation of margin-specific treatment effects. Then, we report additional results.

G.1 Additional details on Mountjoy (2022)'s method

This subsection summarizes the identification and estimation strategy used for the results in Section 7.1. We reproduce many of the arguments from Mountjoy (2022) with our notation and for our setting. Note that in this section, we assume that instruments are treatment-specific (that is, that the instruments induce flows into only one treatment); while this is generally not the case when instruments are judge stringencies, Section 7.1.1 describes our construction of treatment-specific instruments from stringencies.

To begin, we state the "comparable compliers" assumption of Mountjoy (2022) in our notation:

A7. Comparable Compliers (CC)

For all $\tilde{z}_c$ and $\tilde{z}_i$,

$$\begin{aligned}&\lim_{\tilde{z}_c'\uparrow\tilde{z}_c} E[Y(c)|T(\tilde{z}_c',\tilde{z}_i)=c, T(\tilde{z}_c,\tilde{z}_i)=i]\\ =&\lim_{\tilde{z}_i'\downarrow\tilde{z}_i} E[Y(c)|T(\tilde{z}_c,\tilde{z}_i')=c, T(\tilde{z}_c,\tilde{z}_i)=i].\end{aligned}$$

This assumption says that $T=i$ to $T=c$ compliers from decreasing $\tilde{z}_i$ have the same potential outcome when convicted as $T=i$ to $T=c$ compliers from increasing $\tilde{z}_c$ at their limits, where $\tilde{z}_i$ and $\tilde{z}_c$ are the treatment-specific instruments.

Given a treatment-specific instrument for conviction, it is possible to identify a weighted average of LATEs across two margins. In our setting, this is given by:

$$LATE_c = \omega LATE_{d\to c} + (1-\omega) LATE_{i\to c}.$$

This is visualized in Panel (c) of Figure 5, which shows that such variation induces two sets of compliers, those moving from $T=d$ to $T=c$ (in yellow) and those moving from $T=i$ to $T=c$ (in blue).

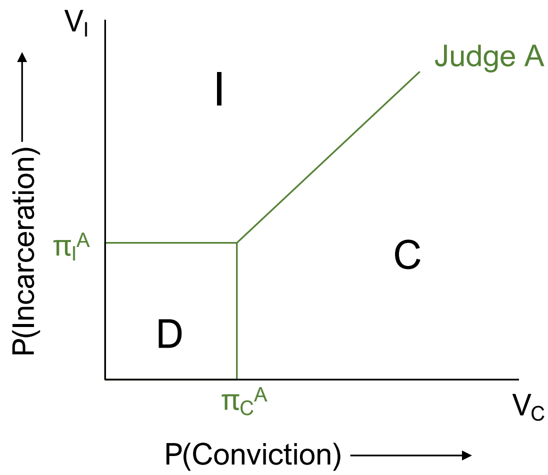
Mountjoy (2022) shows that it is possible to construct the two relevant margin-specific LATEs, as well as their weights, by using variation in two treatment-specific instruments to construct the relevant expected potential outcomes for the two groups. These results involve using local IV and taking limits.

In Figure G.1, we provide some intuition for this result, by considering the effect of incarceration versus conviction. Panel (a) shows treatment choices for Judge A, whose incarceration threshold is π_i^A and conviction threshold is π_c^A. In Panel (b), we consider a second judge, Judge B, who has the same incarceration threshold as Judge A ($\pi_i^A = \pi_i^B$), but a higher conviction threshold ($\pi_c^B > \pi_c^A$). Highlighted in yellow are $T=i$ to $T=c$ compliers: compliers who would be incarcerated under Judge

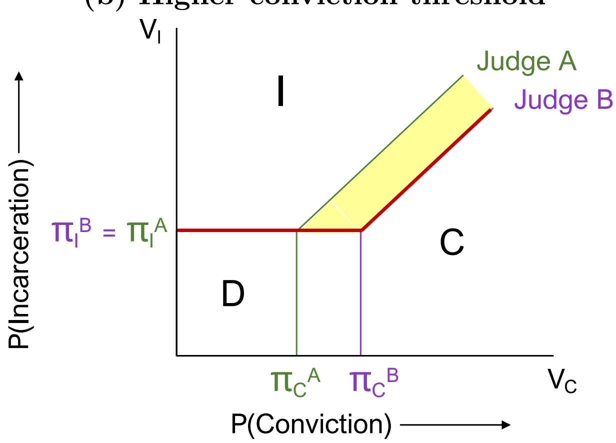
B but convicted under Judge A. We can estimate the recidivism rate for $T = i$ to $T = c$ compliers who get incarcerated, $Y_i^{c\to i}$, by taking the difference in recidivism rates for those incarcerated by Judge A and those incarcerated by Judge B. In Panel (c), we consider a third judge, Judge C. Compared to Judge A, Judge C has the same conviction threshold ($\pi_c^A = \pi_c^C$), but a lower incarceration threshold ($\pi_i^C < \pi_i^A$). Highlighted in yellow are, again, $T = i$ to $T = c$ compliers: compliers who would be convicted under Judge A but incarcerated under Judge C. We can estimate the recidivism rate for $T = i$ to $T = c$ compliers who get convicted, $Y_c^{c\to i}$, by taking the difference in recidivism rates for those convicted by Judge A and those convicted by Judge C. These two instrument shifts – comparing Judge A to Judge B and Judge A to Judge C – generate overlapping $T = i$ to $T = c$ compliers. As we consider smaller changes in π_c and π_i across judges, the difference in areas gets smaller – and differences in mean potential outcomes would isolate the treatment effect of going from conviction to incarceration.

Figure G.1: Illustration of the method in Mountjoy (2022): obtaining $Y_i^{c \to i}$ and $Y_c^{c \to i}$

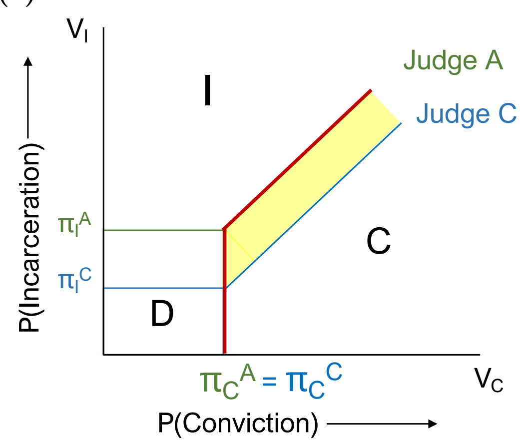
(a) Treatment choices for Judge A



(b) Higher conviction threshold



(c) Lower incarceration threshold



Note: This figure illustrates the insights behind Mountjoy (2022)'s approach. Panel (a) illustrates how a judge divides up the (V_c, V_i) space. Panel (b) illustrates a shift in π_c, holding π_i constant. The area shaded in yellow represents $T = i$ to $T = c$ compliers, who would be incarcerated under Judge B but convicted under Judge A. Panel (c) illustrates a shift in π_i, holding π_c constant. The area shaded in yellow represents the $T = i$ to $T = c$ compliers, who would be incarcerated under Judge C but convicted under Judge A. The areas overlap and the difference in areas gets smaller as we consider smaller changes in π_c and π_i.

In practice, going beyond the illustrative example from Figure G.1, the application focuses on local IV and involves replacing the differences above with derivatives. The identification requires the "comparable compliers" assumption (A7) to hold, which posits that the potential outcome under conviction is the same for marginal compliers between the conviction/incarceration margin regardless of whether the change comes from a marginal increase in $\tilde{z}_c$ or a marginal decrease in $\tilde{z}_i$.

The identification of the relevant parameters involves conditional expectations and partial derivatives. To approximate these conditional expectations and derivatives in estimation, we follow Mountjoy (2022) and assume the conditional expectations are well approximated by a local linear regression centered around the chosen evaluation point of the instruments. The regression includes additive controls for the same variables as the IV regressions described in Section 4, but with coefficients that can vary across evaluation points. We use an Epanechnikov kernel with a bandwidth of 3 and report estimates evaluated at the mean value of the instruments. This approach produces similar estimates when using smaller or larger bandwidths. Inference is based on 200 bootstrap samples. We report 95% confidence intervals based on the bootstraps and significance stars based on the 90%, 95%, and 99% two-sided confidence intervals.

We refer the reader to Mountjoy (2022) for a full discussion of identification and estimation.

G.2 Additional results

Tables G.1 and G.2 provide additional results under alternative assumptions used to construct the treatment-specific instruments. The first set of results comes from assuming a standard multinomial logistic model. While restrictive, this allows for a simple closed form solution for constructing thresholds from shares as explained in the main paper. The second mirrors the mixed model reported in Table 6, but assumes the random effects follow an independent multivariate normal distribution. Confidence intervals for all three approaches are calculated using 200 bootstrap samples.

Overall, the results in Tables G.1 and G.2 are similar in magnitude to Table 6, though the estimates are somewhat larger and tend to be closer to the 2SLS estimates reported in the paper.

Table G.1: Decomposing the impacts of conviction (multinomial logit)

	simple log-ratio				
	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
Panel A: Labeling effect (C vs D)					
Felony Charge:	0.0872	0.182	0.129	0.282*	0.317
	[-0.273,0.944]	[-0.288,1.737]	[-0.514,1.386]	[-0.053,1.797]	[-0.819,3.459]
Felony Conviction:	0.128	0.265*	0.0965	0.420***	0.494**
	[-0.247,0.989]	[-0.046,1.172]	[-0.530,0.911]	[0.131,1.853]	[0.042,4.522]
Felony Incarceration:	0.133	0.0684	-0.0390	0.189	0.224
	[-0.079,1.260]	[-0.233,0.476]	[-0.694,0.810]	[-0.182,1.149]	[-0.249,2.618]
Panel B: Decarceration (C vs I)					
Felony Charge:	0.107***	-0.0695	-0.0170	0.00264	0.0148
	[0.049,0.187]	[-0.213,0.054]	[-0.135,0.120]	[-0.161,0.133]	[-0.166,0.206]
Felony Conviction:	0.116***	-0.0570	-0.0349	0.0351	0.0327
	[0.057,0.199]	[-0.180,0.073]	[-0.162,0.094]	[-0.097,0.187]	[-0.135,0.229]
Felony Incarceration:	0.0820***	-0.0240	-0.0568	0.0147	0.00493
	[0.031,0.158]	[-0.117,0.082]	[-0.162,0.027]	[-0.113,0.132]	[-0.127,0.166]
Panel C: Net Effect					
Felony Charge:	0.103***	-0.0224	0.00915	0.0550	0.0690
	[0.051,0.188]	[-0.122,0.097]	[-0.088,0.131]	[-0.064,0.178]	[-0.061,0.260]
Felony Conviction:	0.118***	0.00335	-0.0114	0.107**	0.115*
	[0.082,0.202]	[-0.085,0.137]	[-0.117,0.102]	[0.006,0.266]	[-0.008,0.282]
Felony Incarceration:	0.0906***	-0.00648	-0.0536	0.0473	0.0441
	[0.048,0.166]	[-0.098,0.099]	[-0.142,0.020]	[-0.050,0.167]	[-0.062,0.199]
Controls	Yes	Yes	Yes	Yes	Yes

Note: This table decomposes the local IV estimates (Panel C) into the "Labeling effect" (C vs D) and "Decarceration" (C vs I) using an unordered multinomial model, based on the methodology developed in Mountjoy (2022). 95% confidence intervals are reported in brackets and are based on 200 bootstrap samples. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01. In this table, judge-specific latent preferences are calculated under the stronger assumption that case outcomes are determined by a multinomial logit.

Table G.2: Decomposing the impacts of conviction (independent mixed logit)

	mixed logit with independent normal random effects				
	Year 1	Year 2-4	Year 5-7	Year 1-4	Year 1-7
Panel A: Labeling effect (C vs D)					
Felony Charge:	0.0440	0.105*	0.0946*	0.130*	0.141**
	[-0.027,0.129]	[-0.007,0.214]	[-0.001,0.185]	[-0.003,0.259]	[0.010,0.276]
Felony Conviction:	0.0609*	0.138***	0.0767*	0.175***	0.197***
	[-0.004,0.139]	[0.042,0.254]	[-0.013,0.164]	[0.057,0.309]	[0.086,0.343]
Felony Incarceration:	0.0346	0.0856**	0.0328	0.0902	0.112*
	[-0.024,0.103]	[0.000,0.192]	[-0.044,0.113]	[-0.026,0.191]	[-0.031,0.240]
Panel B: Decarceration (C vs I)					
Felony Charge:	0.0392***	-0.0379	-0.00397	-0.0181	0.0100
	[0.013,0.070]	[-0.101,0.013]	[-0.066,0.065]	[-0.085,0.056]	[-0.083,0.107]
Felony Conviction:	0.0325**	-0.0308	-0.0166	-0.00939	0.00621
	[0.006,0.061]	[-0.094,0.014]	[-0.071,0.030]	[-0.073,0.063]	[-0.089,0.104]
Felony Incarceration:	0.0117	-0.0387*	-0.0191	-0.0279	0.0162
	[-0.015,0.035]	[-0.091,0.009]	[-0.071,0.030]	[-0.095,0.035]	[-0.066,0.113]
Panel C: Net Effect					
Felony Charge:	0.0403***	-0.00383	0.0212	0.0173	0.0433
	[0.018,0.068]	[-0.044,0.046]	[-0.028,0.074]	[-0.037,0.078]	[-0.025,0.122]
Felony Conviction:	0.0391***	0.00946	0.00716	0.0346	0.0547
	[0.019,0.065]	[-0.044,0.052]	[-0.038,0.053]	[-0.011,0.089]	[-0.014,0.129]
Felony Incarceration:	0.0170	-0.00907	-0.00590	0.000260	0.0406
	[-0.007,0.038]	[-0.048,0.031]	[-0.048,0.038]	[-0.053,0.049]	[-0.022,0.108]
Controls	Yes	Yes	Yes	Yes	Yes

Note: This table decomposes the local IV estimates (Panel C) into the "Labeling effect" (C vs D) and "Decarceration" (C vs I) using an unordered multinomial model, based on the methodology developed in Mountjoy (2022). 95% confidence intervals are reported in brackets and are based on 200 bootstrap samples. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01. This table computes the judge-specific latent preferences using a mixed logit model where there is a random effect on the intercepts. The random effect is assumed to be distributed with a multivariate normal distribution with zeros in the off diagonal terms of the covariance matrix.

H Impacts of incarceration: additional evidence from sentencing guidelines

In this Appendix, we provide more details on the empirical approach followed in our regression discontinuity design analyses from Section 7.2.

H.1 Empirical setup

Calculating the sentencing score. The Virginia sentence guidelines were developed in the 1980s to harmonize practices across judges and reduce disparities across similar defendants (Farrar-Owens, 2013). Information on the sentence guidelines is available to all parties during negotiations.

The diagram in Figure H.1 describes the order in which the different sentencing worksheets are filled out. The first worksheet determines whether a person convicted of a felony is recommended for prison (more than one year of incarceration). This worksheet, called "Worksheet A", consists of a series of questions pertaining to the offense and criminal history. Each question has a number of points associated with it; the sum of these points is the "incarceration-length score". Those who score above a cutoff are recommended for prison. Those who score below the cutoff are recommended for probation or jail, where recommended jail sentences are under a year in length.

Based on the cutoff in Worksheet A, either Worksheet B (for those below the cutoff) or Worksheet C (for those above the cutoff) is used to calculate the final guidelines-recommended sentence. Worksheet B also has a discontinuity that is useful for our analysis. Defendants who score above a particular cutoff on the 'probation/jail score' are recommended for short jail sentences, while defendants who score below that cutoff are recommended for probation.

Offenses are sorted into 16 offense categories and each category has a slightly different worksheet. The worksheets are filled out by a probation officer or a prosecutor and then given to a judge during sentencing. The worksheet package contains a cover sheet, which has a summary of information related to the case. The guidelines-recommended sentence and range is displayed prominently on the cover sheet. An example of Worksheet A can be found in Appendix H.5; the other worksheets follow a similar organization.

Empirical approach. To conduct this analysis, we compare people who score just below and just above our worksheet thresholds. The main assumption for this to yield causal estimates of the effects of tougher sentences is that potential outcomes are smooth across the cutoff. This might not hold if, for example, legal actors are able to manipulate the scores. Three institutional details in our setting help mitigate this concern. First, the sentence guidelines are discretionary, not binding. Thus it is not be necessary for legal actors to manipulate the score to achieve a certain sentence. Second, legal actors may pay more attention to the final recommended sentence as calculated on Worksheet B or Worksheet C, rather than the intermediary score calculated on Worksheet A. Therefore, concerns of manipulation on the incarceration-length score (derived from Worksheet A) might not be as strong, simply because it's less salient. Third, from the legislator's standpoint, the goal of these worksheets was to reduce unjustified disparities. It therefore seems unlikely that the sharp sentencing discontinuities observed at the cutoff in the incarceration-length score were created on purpose. In Section H.2 below, we provide evidence that there is no change in characteristics at the cutoff, along with tests for bunching in the running variable on either side of the cutoff.

An additional challenge in our setting is that the running variable is discrete, generating difficulties in estimating accurate confidence intervals. To address this, we adopt the technique developed by Kolesár and Rothe (2018) – "K&R" henceforth – designed specifically for regression discontinuity with a discrete running variable. As in other RD settings, we want to estimate a function of the form:

$$y_{i,s} = \beta * \mathbb{1}(s \geq 0) + f(s) * \mathbb{1}(s \geq 0) + g(s) * \mathbb{1}(s < 0) + \epsilon \tag{14}$$

where $y_{i,s}$ is the outcome of person in case i having obtained a sentencing score of s.[71] Our main coefficient of interest is β. The challenge is to estimate the form of $f(.)$ and $g(.)$, especially close to the cutoff.

Typical approaches in RD consist of fitting specifications on either side of the cutoff. However, these approaches assume that bias can be minimized by reducing the bandwidth. In the discrete setting, the bandwidth cannot asymptotically go to zero, because there are no observations in between each discrete bin. The scarcity of points close to the cutoff could lead to misspecification error: in the absence of additional assumptions, it is unclear what the behavior of the functions of interest would be close to the cutoff, resulting in misspecified confidence intervals.

K&R offer an approach to determine confidence intervals, by estimating plausible behaviors of the potential outcome function close to the cutoff based on its behavior at other points. By fitting a linear regression through points at the left and right of the cutoff, we might be missing non-linearities closer to the cutoff. We cannot use observations "very close" to the cutoff to estimate this, since the discrete nature of the score hinders the credibility of limit arguments. K&R determine credible bounds for the second derivatives of $f(.)$ and $g(.)$ close to the cutoff, based on its behavior further from the cutoff, to estimate the magnitude of plausible deviations from the linear estimation. We need to choose a parameter K which is the upper bound of the absolute value of second derivative of the conditional expectation function. This tells us how quickly the functions $f(.)$ and $g(.)$ can change. Using K, we can construct confidence intervals that reflect how far away from the linear approximation the true conditional expectation function might be based on its expected behavior at other points.

To choose K, we follow the approach developed by Imbens and Wager (2019) and implemented by Goldsmith-Pinkham et al. (2021). We take a large window of nine points to the left of the cutoff and fit a quadratic function of the sentencing score to the data.[72] We take the coefficient on the quadratic term, take the absolute value, and multiply it by four. Intuitively, this means that we allow the rate of change (2nd derivative) of $f(.)$ at the cutoff to be two times that of the estimated rate of change between -9 and -1 from a second order polynomial. When we estimate the optimal bandwidth, we obtain an optimal choice of equal to or close to 5 for many of our main outcomes. In order to keep bandwidths constant across outcomes and time periods, we use a bandwidth of 5 in all specifications.

[71] As a reminder, the sentencing score is either the incarceration-length score or the probation/jail score.

[72] We focus on the left of the cutoff, since we have more observations there.

H.2 Additional results: balance and marginal cases

Balance tests. Figure H.2 (H.3) and Table H.2 (H.3) present balance tests for the intensive margin experiment based on Worksheet A (extensive margin experiment, Worksheet B). We first iperform analyses of defendant characteristics, such as demographics or criminal history, and find no notable discontinuities. We then turn to legal actor decisions. Since inputs to the worksheets and how they translate into sentences is common knowledge, it is possible that some savvy legal actors might try to manipulate inputs. For example, a better defense attorney might push harder to drop certain charges if their client has a score close the cutoff, in order to push them just below the cutoff and avoid longer recommended sentences. If defense attorneys were trying to push their clients to the left of the cutoff, one way this could manifest is by having more charges dropped just before the cutoff. That is because some of the points are linked to number of offenses for which a person is convicted. This does not seem to be happening. We also look at measures of defendant poverty, which can affect quality of representation (Agan et al., 2021b).[73] We do not find evidence of a discontinuity at the cutoff, suggesting that quality of representation does not change at this point.

We do find one difference: defendants in the incarceration-length sample are about 2.8 percentage points more likely to have their case resolved by plea just before than just after the cutoff (Panel B of table H.2). This could be because the longer sentences offered to those just above the threshold make people more willing to "risk it" in court. Since taking the case to trial increases the likelihood of dismissal by 10 percentage points, a 2.8 percentage point increase in the trial rate would lead to losing 0.28% of the sample right above the threshold. Given how small the differences in conviction is at the threshold, and the fact that we see no detectable differences in observable characteristics, we think that this is unlikely to affect our research design too much. We also note that we do not find this discontinuity for the probation/jail sample, so these concerns do not apply to that set of analyses.

Lastly, we examine the distribution of the running variables to evaluate whether there is excess mass right above or below the cutoff. Such excess mass would be consistent with strategic manipulation of the scores to nudge defendants above or below the discontinuity in guidelines-recommended sentence. These analyses are shown in Figures H.2 (a) and H.3 (a) for the incarceration-length score and the probation/jail score, respectively. Visual inspection reveals possible excess mass below the cutoff for the incarceration-length score. Though, the distribution is not smooth, making it hard to infer whether this bunching is just a natural byproduct of a lumpy distribution, or the result of strategic manipulation. There is no visible bunching around the cutoff for the probation/jail score.

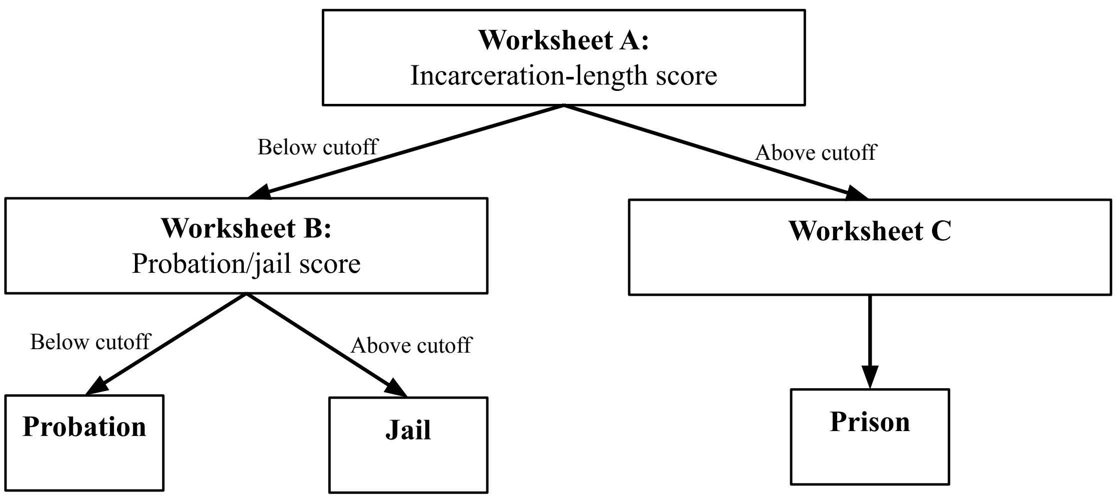
Marginal case. Appendix Table H.5 compares the characteristics of marginal cases to those of the full sample in the relevant experiment, where marginal cases are defined as those scoring right below or right above the cutoff. The biggest difference between marginal cases and the full sample for Worksheet A is that marginal cases are much more likely to have prior incarceration: 87% had been incarcerated in the past, compared to 65% for the sample overall. This set aside, marginal cases are similar across

[73] We proxy poverty by whether a defendant comes from ZIP codes where the percent of people reporting less than $25,000 (less than $10,000) per year to the IRS was above the median within our sample.

offenses, but tend to be slightly younger. For worksheet B, there are differences across offense types: people convicted of a drug offense are more likely to be moved by the policy, while people convicted with property crimes are less so. Marginal cases are also more likely to have been incarcerated in the past (65% compared to 54%). Note that the marginal cases in the RD and IV experiments are different (as an example, 21% of the IV incarceration marginal cases had a prior felony conviction in the last 5 years, compared to 85% of the RD marginal cases). Yet, our results are similar across both experiments, suggesting that the differences in composition are not yielding different findings.

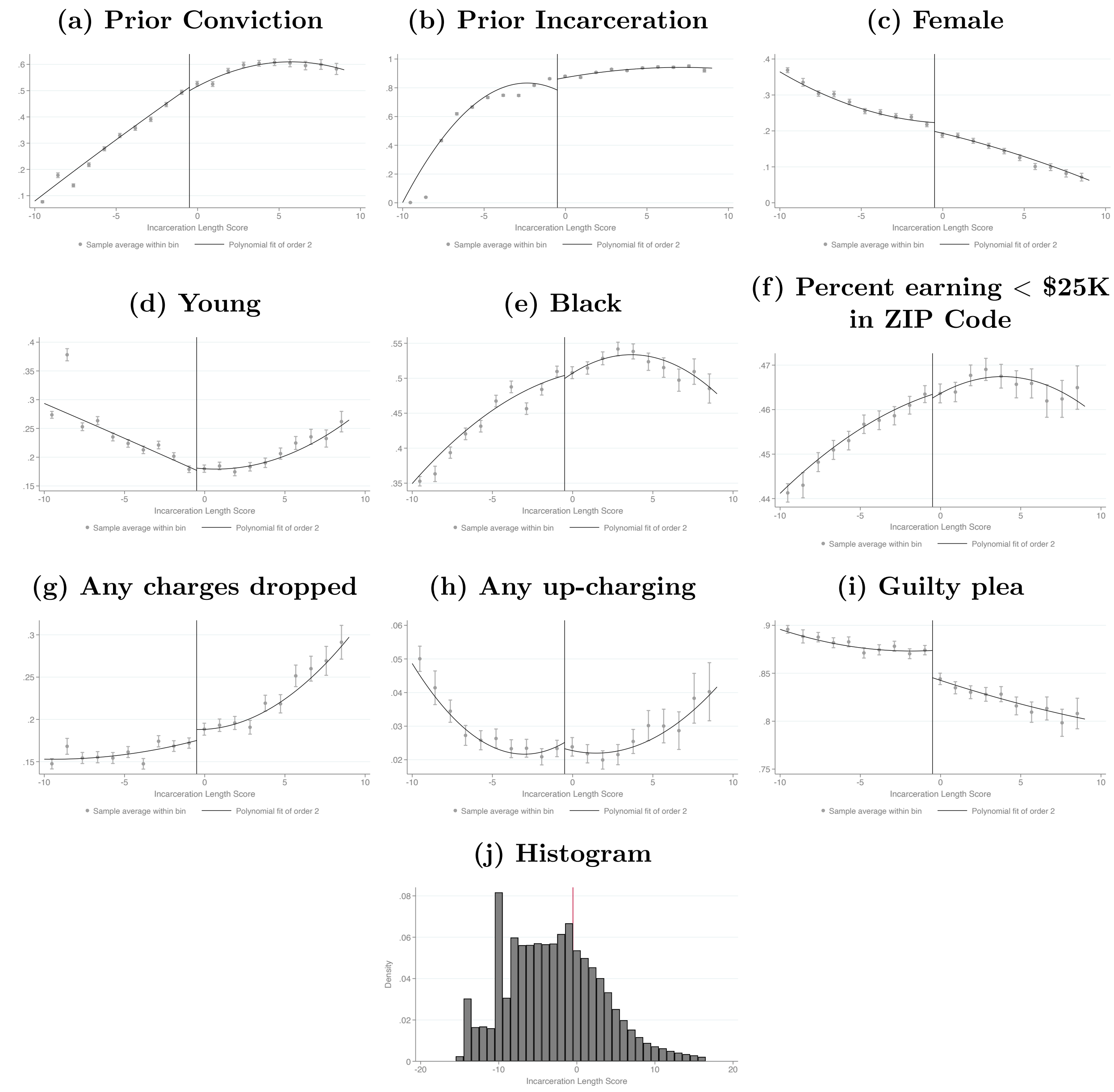
H.3 Appendix Figures: RD analyses

Figure H.1: Flowchart of felony sentencing determination in Virginia



Note: This figure presents a flowchart describing the sentencing process in Virginia after a felony conviction, and how and when different Worksheets are used.

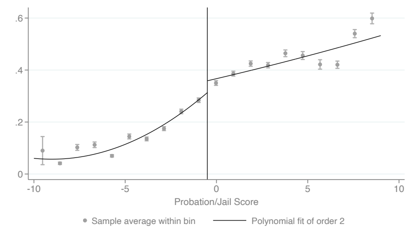
Figure H.2: Balance tests – incarceration-length sample



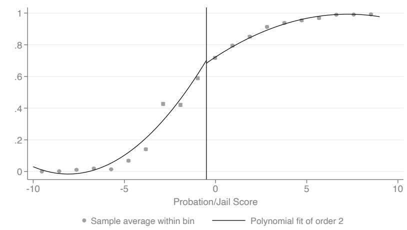
Note: Panels (a) - (i) show RD plots for various demographic variables and case characteristics. Panel (j) shows the distribution of incarceration-length scores around the cutoff. The incarceration-length score is normalized so that the cutoff is at zero.

Figure H.3: Balance tests – probation/jail sample

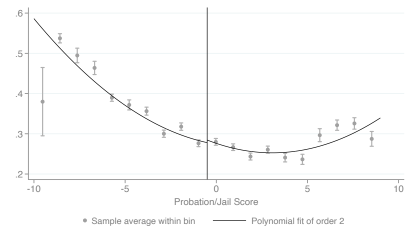
(a) Prior Conviction



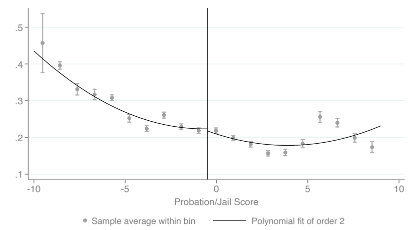
(b) Prior Incarceration



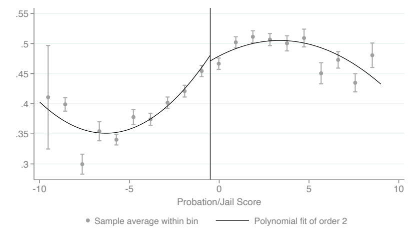
(c) Female



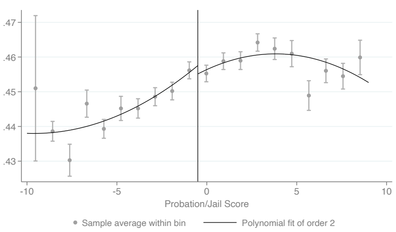
(d) Young



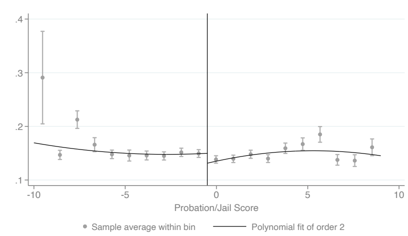
(e) Black



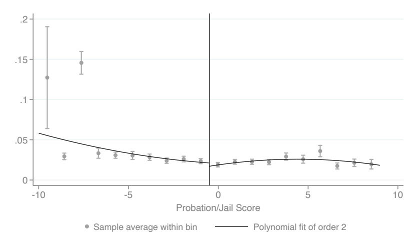
(f) Percent earning < $25K in ZIP Code



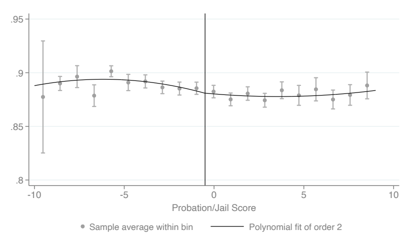
(g) Any charges dropped



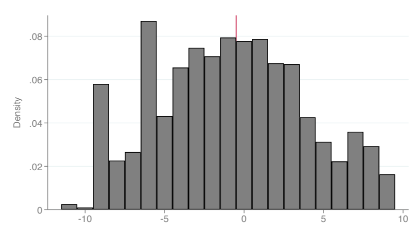
(h) Any up-charging



(i) Guilty plea



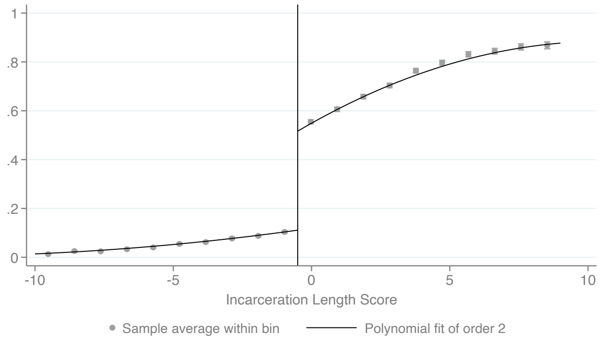
(j) Histogram



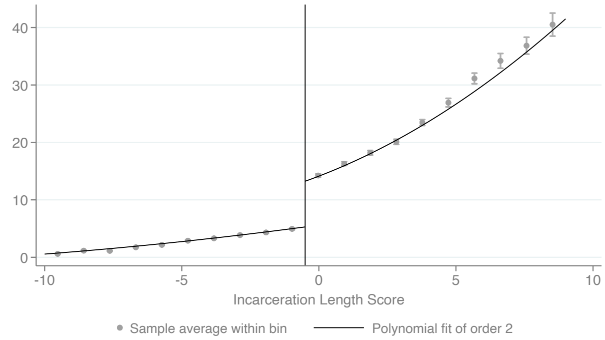
Note: Panels (a) - (i) show RD plots for various demographic variables and case characteristics. Panel (j) shows the distribution of probation/jail scores around the cutoff. The probation/jail score is normalized so that the cutoff is at zero.

Figure H.4: RD First stage and outcome graphs – incarceration-length sample

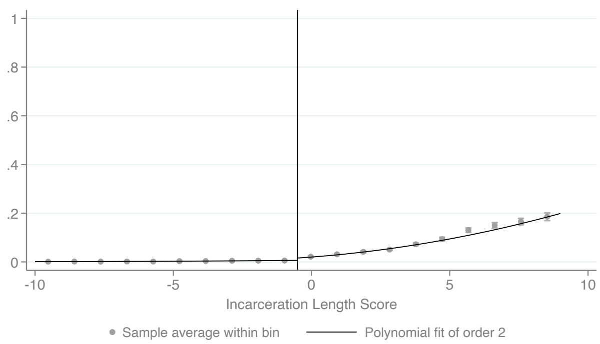
(a) Incarcerated for at least 1 year



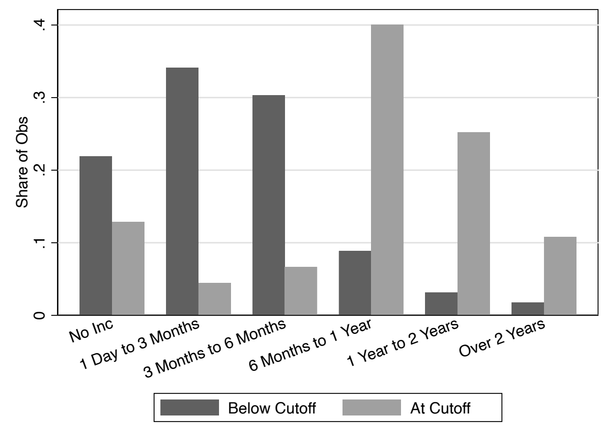
(b) Sentence in months



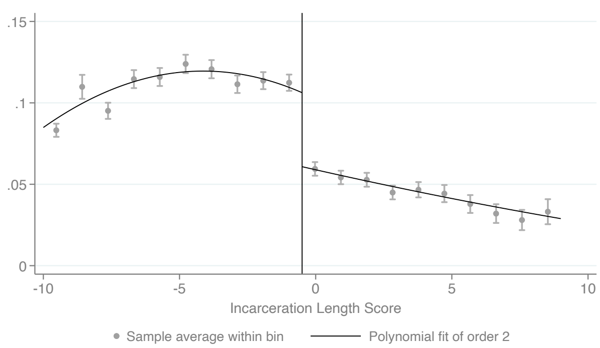
(c) Incarcerated for at least five years



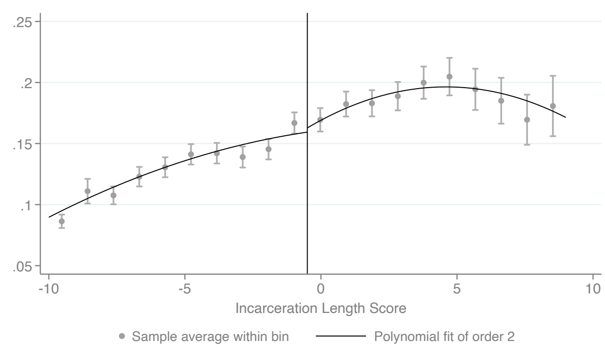
(d) Sentencing distributions for those right above/below the cutoff



(e) New Charge within 1 Year of Sentencing

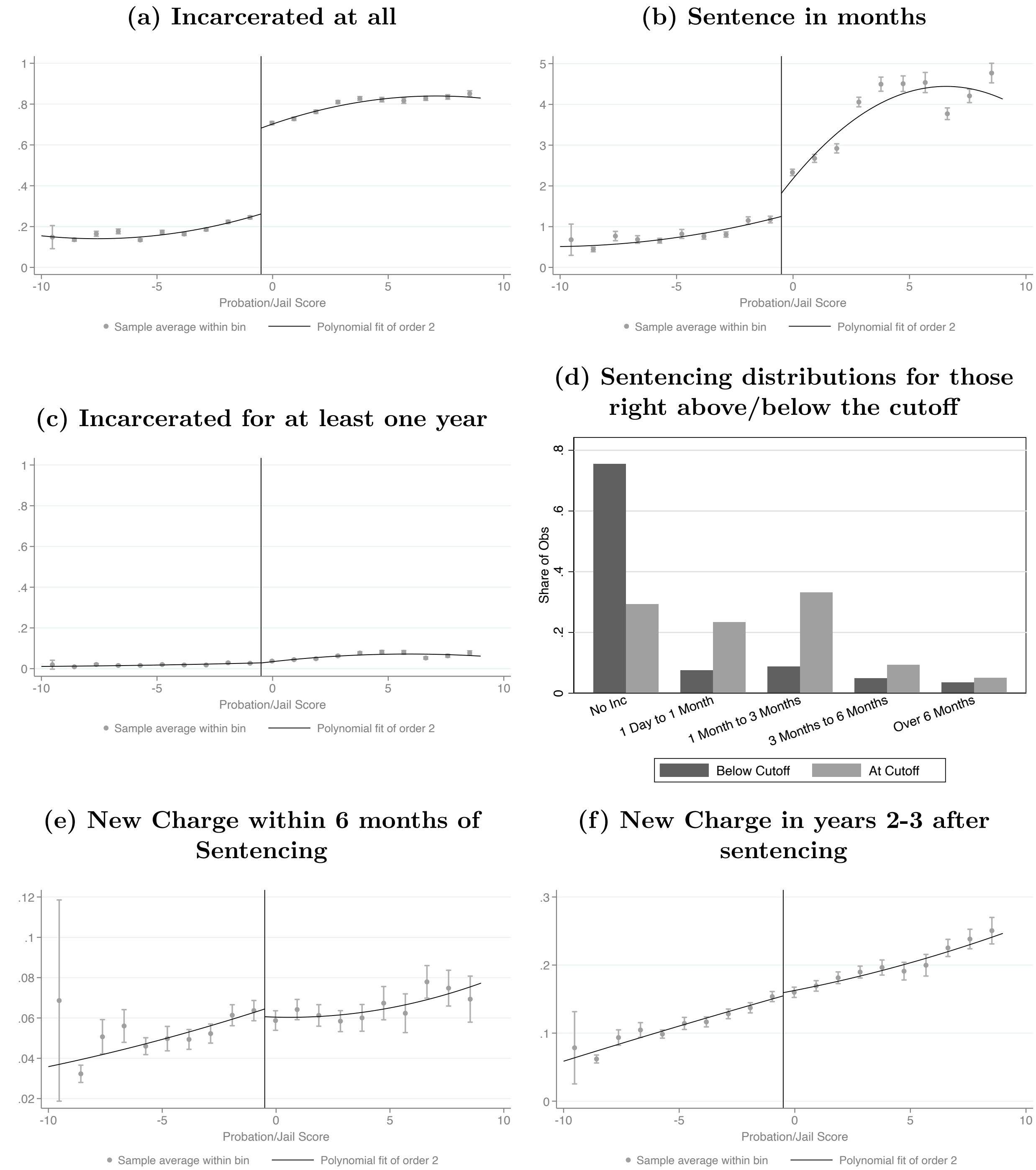


(f) New Charge in years 8-10 after sentencing



Note: Panel (a) shows the RD plot for being incarcerated for at least one year around the discontinuity in the incarceration-length score. Panel (b) shows the same plot for months sentenced and panel (c) shows the same plot for being sentenced to at least five years. Panel (d) shows the distribution of sentence lengths for those just above and just below the cutoff. Panel (e) shows RD plots for recidivism-defined as a binary variable for having at least one new charge one year post sentencing and panel (f) shows recidivism within 8-10 years post sentencing.

Figure H.5: RD First stage and outcome graphs – probation/jail score



Note: Panel (a) shows the RD plot for being incarcerated at all. Panel (b) shows the same plot for months sentenced and panel (c) shows the same plot for being sentenced to at least one year. Panel (d) shows the distribution of sentence lengths for those just above and just below the cutoff. Panel (e) shows RD plots for recidivism-defined as a binary variable for having at least one new charge six months post sentencing and panel (f) shows recidivism within 2-3 years post sentencing.

H.4 Appendix Tables: RD analyses

Table H.1: Summary statistics: RD sample

	Incarceration length worksheet	Probation/jail worksheet
	mean	mean
Offenses		
Assault	0.05	0.00
Burglary	0.10	0.00
Drug	0.41	0.58
Larceny	0.37	0.41
Miscellaneous	0.01	0.01
Robbery	0.02	0.00
Sexual assault	0.03	0.00
Defendant characteristics		
Black	0.46	0.44
Female	0.25	0.32
Under 23	0.23	0.23
% of ppl in zip earning <25K	0.46	0.45
Incarceration		
Recommended for prison	0.35	0.00
Prior incarceration	0.65	0.54
Prior circuit crt. felony convic.	0.38	0.29
Carceral sentence	0.62	0.47
Jail sentence	0.35	0.45
Prison sentence	0.28	0.04
Sentence >= 5 years	0.03	0.00
Months of sentence	10.39	2.13
Post-release		
New felony charge within 1 year	0.09	0.11
Observations	230,357	130,692

Note: This table shows means of relevant variables for the incarceration-length sample from Worksheet A and the probation/jail sample from Worksheet B.

Table H.2: Balance: RD estimates for incarceration-length sample

	(1) In Virginia 5-7yrs	(2) Any Prior Chrg.	(3) Prior Incarc.	(4) Female	(5) Young	(6) Black
Panel A: demographic balance						
RD Estimate:	-0.004	-0.007	-0.011	-0.020	0.004	-0.007
	[-0.016,0.009]	[-0.019,0.006]	[-0.171,0.149]	[-0.037,-0.003]	[-0.028,0.035]	[-0.034,0.019]
N	144,761	125,610	144,761	134,450	144,117	134,446
Control mean	0.77	0.41	0.78	0.24	0.21	0.48
	Plea	Dropped Chrg.	Upgrade Chrg.	Zip <10K	Zip <25K	
Panel B: income & legal actor balance						
RD Estimate:	-0.028	0.012	0.001	-0.001	-0.002	
	[-0.039,-0.016]	[-0.006,0.029]	[-0.007,0.010]	[-0.002,0.001]	[-0.006,0.003]	
N	144,761	120,223	120,223	94,546	94,546	
Control mean	0.87	0.17	0.02	0.19	0.46	

Note: Panel A shows RD estimates of discontinuities in various predetermined characteristics across the cutoff in the incarceration-length score. Panel B tests for discontinuities at the cutoff in outcomes of the criminal proceedings, such as whether the case resolved in a guilty plea, whether there were any dropped charges, whether there were any charges that were upgraded from misdemeanor to felony, and various measures of indigency within the defendant's ZIP Code. Below the estimates, we present in brackets confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Section 7.2 for a discussion of parameter choices.

Table H.3: Balance: RD estimates for the probation/jail sample

	(1) In Virginia 5-7yrs	(2) Any Prior Chrg.	(3) Prior Incarc.	(4) Female	(5) Young	(6) Black
Panel A: demographic balance						
RD Estimate:	-0.004	0.033	0.044	0.013	0.010	0.004
	[-0.033,0.024]	[-0.016,0.082]	[-0.098,0.185]	[-0.026,0.052]	[-0.026,0.046]	[-0.049,0.057]
N	105,839	91,271	105,839	97,089	105,289	97,085
Control mean	0.77	0.20	0.36	0.32	0.24	0.41
	Plea	Dropped Chrg.	Upgrade Chrg.	Zip <10K	Zip <25K	
Panel B: zip income & legal actor balance						
RD Estimate:	-0.003	-0.014	-0.004	-0.001	-0.003	
	[-0.015,0.010]	[-0.028,-0.001]	[-0.010,0.002]	[-0.003,0.001]	[-0.009,0.003]	
N	105,839	83,008	83,008	65,263	65,263	
Control mean	0.89	0.15	0.03	0.18	0.45	

Note: Panel A shows RD estimates of discontinuities in various predetermined characteristics at the cutoff in the probation/jail score. Panel B tests for discontinuities across the cutoff in outcomes of the criminal proceedings, such as whether the case resolved in a guilty plea, whether there were any dropped charges, whether there were any charges that were upgraded from misdemeanor to felony, and various measures of indigency within the defendant's ZIP Code. Below the estimates, we present in brackets confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Section 7.2 for a discussion of parameter choices.

Table H.4: First stage-RD estimates

	Months	Any Incar	Incar > 1 yr	Incar > 5 yrs
Panel A: Incarceration-length sample				
RD Estimate:	8.005	0.0652	0.420	0.00964
	[7.569,8.442]	[-0.017,0.147]	[0.404,0.437]	[0.007,0.013]
N	144,761	144,761	144,761	144,761
Control mean	3.91	0.70	0.08	0.00
Panel B: Probation/jail sample				
RD Estimate:	0.726	0.429	0.00308	-0.000585
	[0.532,0.919]	[0.393,0.465]	[-0.003,0.009]	[-0.002,0.000]
N	105,839	105,839	105,839	105,839
Control mean	0.96	0.20	0.02	0.00
Panel C: Probation/jail samp. no prior inc.				
RD Estimate:	0.820	0.431	0.00630	-0.00103
	[0.200,1.440]	[0.350,0.512]	[-0.014,0.027]	[-0.005,0.003]
N	8,870	8,870	8,870	8,870
Control mean	0.81	0.18	0.02	0.00

Note: This table shows RD estimates of how cutoffs in the incarceration-length score and probation/jail score affect incarceration length (column 1), the probability of getting a carceral sentence (column 2), and the probability of having a sentence longer than 1 year (column 3) or 5 years (column 4). Below the estimates, we present in brackets confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Section 7.2 for a discussion of parameter choices.

Table H.5: Marginal cases in the RD study

	Incarceration length worksheet		Probation/jail worksheet	
	P(X=x)	P(X=x| Marginal)	P(X=x)	P(X=x| Marginal)
Prior Conviction	0.636	0.852	0.521	0.565
	(0.481)	(0.355)	(0.500)	(0.496)
Female	0.245	0.204	0.320	0.277
	(0.430)	(0.403)	(0.466)	(0.447)
Black	0.458	0.507	0.438	0.459
	(0.498)	(0.500)	(0.496)	(0.498)
Prior incarceration	0.651	0.871	0.535	0.651
	(0.477)	(0.335)	(0.499)	(0.477)
Drugs	0.412	0.393	0.576	0.815
	(0.492)	(0.488)	(0.494)	(0.388)
Property	0.496	0.491	0.413	0.173
	(0.500)	(0.500)	(0.492)	(0.378)
Violent	0.073	0.098	0.000	0.000
	(0.260)	(0.298)	(0.000)	(0.000)
Other	0.040	0.047	0.011	0.012
	(0.197)	(0.212)	(0.105)	(0.111)
Observations	230357	27556	152663	20609

Note: This table compares socio-demographic characteristics of compliers to that of the full sample for the RD sample.

H.5 Example of sentencing worksheet

Drug/Schedule I/II ❖ **Section A** **Offender Name:** ____________________

◆ **Primary Offense**

A. Possess Schedule I or II drug

	1 count	1
	2 counts	3
	3 counts	8

B. Sell, Distribute, Possession with Intent Schedule I or II drug

	1 count	12
	2 counts	13
	3 counts	14
	4 counts	15

C. Sell, etc. Schedule I, II drug to minor (1 count) 11

D. Accommodation - Sell, Distribute, Possession with Intent Schedule I or II drug

	1 count	5
	2 counts	7

E. Sell, etc. imitation Schedule I or II drug (1 count) 4

Score [|]

◆ **Primary Offense Additional Counts** Total the maximum penalties for counts of the primary not scored above → [0 |]

Years:				
	5 - 10	1	31 - 42	4
	11 - 21	2	43 or more	5
	22 - 30	3		

◆ **Additional Offenses** Total the maximum penalties for additional offenses, including counts → [0 |]

Years:				
	Less than 4	0	22 - 30	3
	4 - 10	1	31 - 42	4
	11 - 21	2	43 or more	5

◆ **Knife or Firearm in Possession at Time of Offense** — **If YES, add 2** → [0 |]

◆ **Conviction in Current Event Requiring Mandatory Minimum Term (6 mos or more)** **If YES, add 9** → [0 |]

◆ **Mandatory Firearm Conviction for Current Event** — **If YES, add 7** → [0 |]

◆ **Prior Convictions/Adjudications** Total the maximum penalties for the 5 most recent and serious prior record events → [0 |]

Years:		
	Less than 7	0
	7 - 26	1
	27 - 48	2
	49 or more	3

◆ **Prior Incarcerations/Commitments** — **If YES, add 2** → [0 |]

◆ **Prior Felony Drug Convictions/Adjudications** → [0 |]

Number:		
	1 - 2	1
	3 - 4	2
	5	3
	6 or more	4

◆ **Prior Juvenile Record** — **If YES, add 1** → [0 |]

◆ **Legally Restrained at Time of Offense** → [0 |]

None	0
Other than parole/post-release, supervised probation or CCCA	1
Parole/post-release, supervised probation or CCCA	4

SCORE THE FOLLOWING FACTOR **ONLY** IF PRIMARY OFFENSE IS **POSSESSION OF SCHEDULE I/II DRUG** (§ 18.2-250(A,a))

◆ **Two or More Prior Felony Convictions/Adjudications** — **If YES, add 2** → [0 |]
For Possession, Possession with Intent, Distribution, Manufacture or Sale of Schedule I or II Drug

Total Score → [|]
If total is 10 or less, go to **Section B**. If total is 11 or more, go to **Section C**.

Drug Schedule I or II/ Section A Eff. 7-1-09