NBER WORKING PAPER SERIES

CONVICTION, INCARCERATION, AND RECIDIVISM:
UNDERSTANDING THE REVOLVING DOOR

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

Conviction, Incarceration, and Recidivism: Understanding the Revolving Door
John Eric Humphries, Aurelie Ouss, Kamelia Stavreva, Megan T. Stevenson, and Winnie
van Dijk
NBER Working Paper No. 32894
August 2024, Revised March 2025
JEL No. J0, K4

ABSTRACT

Noncarceral conviction is a common outcome of criminal court cases: for every individual incarcerated, there are approximately three who were recently convicted but not sentenced to prison or jail. We extend the binary-treatment judge IV framework to settings with multiple treatments and use it to study the consequences of noncarceral conviction. We outline assumptions under which widely-used 2SLS regressions recover margin-specific treatment effects, relate these assumptions to models of judge decision-making, and derive an expression that provides intuition about the direction and magnitude of asymptotic bias when a key assumption on judge decision-making is not met. We find that noncarceral conviction (relative to dismissal) leads to a large and long-lasting increase in recidivism for felony defendants in Virginia. In contrast, incarceration (relative to noncarceral conviction) leads to a short-run reduction in recidivism, consistent with incapacitation. Our empirical results suggest that noncarceral felony conviction is an important and overlooked driver of recidivism.

John Eric Humphries
Yale University
and NBER
johneric.humphries@yale.edu

Aurelie Ouss
Department of Criminology
University of Pennsylvania
571 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104
and NBER
aouss@upenn.edu

Kamelia Stavreva
Columbia University
kes2220@columbia.edu

Megan T. Stevenson
University of Virginia
mstevenson@law.virginia.edu

Winnie van Dijk
Department of Economics
Yale University
87 Trumbull Street, B228
New Haven, CT 06520
and NBER
winnie.vandijk@yale.edu

I 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 they are driven by external factors such as addiction, mental health, neighborhood disadvantage, or limited labor market opportunities. Much of the available quantitative research has focused on how *incarceration* affects recidivism. However, *noncarceral conviction* (a conviction that does not result in incarceration) is a frequent outcome in the criminal court system.[2] For instance, in 2010, 2.7 individuals were on probation for every person who was incarcerated (Phelps, 2013). A noncarceral conviction could directly affect recidivism through several channels. It may induce crime by reducing its opportunity cost. For example, a conviction record could make it harder to find employment, making crime relatively more attractive. A conviction could also 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 recent conviction on their 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.

In this paper, we provide new evidence on how both felony noncarceral conviction and incarceration affect future criminal justice involvement. Our main approach follows existing research by using quasi-random assignment of cases to judges as a source of exogenous variation, but our discussion formalizes an extension of this research design from the binary-treatment case to the multiple-treatment case. Our goal is to learn about *margin-specific* treatment effects: causal impacts of noncarceral conviction relative to dismissal of all charges, and causal impacts of incarceration relative to noncarceral conviction. These quantities allow us to isolate the role of mechanisms that come into play when someone is convicted without a carceral sentence (such as the mark of a criminal record) from the role 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, new convictions, and new carceral sentences. Following the literature, we henceforth refer to these outcomes as "recidivism." Our results point to noncarceral conviction as an important, long-lasting driver of recidivism, consistent with a criminogenic effect of a felony conviction record. In contrast, we find that incarceration leads to only a short-term decline in recidivism, consistent with incapacitation.

Our discussion proceeds in three parts. First, we develop an empirical framework for interpreting judge-stringency 2SLS estimands in a multiple-treatment setting with full treatment effect heterogeneity. Prior applied work using 2SLS with multiple treatments has often

---

[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, Durose, and Markman, 2018).

[2]We will at times refer to "noncarceral conviction" as "conviction" for brevity.

relied on instruments that are reasonably thought of as varying the net payoff to taking up a "focal" treatment (e.g., Kline and Walters, 2016; Kirkeboen, Leuven, and Mogstad, 2016; Mountjoy, 2022). For such instruments, it may be justifiable to assume that they are *treatment-specific*, i.e., they either encourage or discourage take-up of the focal treatment and do not cause any switches between other "non-focal" treatments. This property, combined with the usual IV assumptions, ensures that the estimand from a standard 2SLS regression identifies a causal effect of the focal treatment, relative to a mixture of alternatives.[3] However, judge stringency instruments vary the shares of cases that are allocated to particular court outcomes. Therefore, they cannot generally be thought of as varying the net payoff to taking up a particular focal treatment, as in the examples cited above.

We argue that this property of judge stringency instruments implies that insights from prior work do not directly carry over. On the one hand, we show that treatment specificity is sufficient for 2SLS with judge stringency instruments to identify *margin-specific* causal effects, unlike in the previously-cited applications. On the other hand, requiring stringency instruments to be treatment-specific could be considered a strong restriction on judge behavior, while it was considered reasonable for other types of instruments. We provide intuition for the restrictiveness of this assumption by examining how it constrains models of judge decision-making. We consider three commonly-used discrete-choice models, applied to judge decision-making over three court outcomes: dismissal, noncarceral conviction, and incarceration. Specifically, we consider ordered, sequential, and multinomial choice models. The judge-stringency instruments are treatment-specific only in the ordered model. For the sequential and unordered models, which are more realistic in our setting, at least one of the instruments is not treatment-specific. However, all three choice models satisfy a weaker assumption that we label conditional pairwise monotonicity (CPM). This assumption, related to the "no defiers" assumption from the binary case, states that an instrument induces flows in only one direction across each margin.

We then derive an expression for the asymptotic bias in the 2SLS estimand under CPM. The bias term is additive and easy to interpret. It provides intuition about the direction and magnitude of asymptotic bias when CPM holds, but treatment specificity does not. Moreover, it clarifies how restrictions on treatment effect heterogeneity, or on the relative effects for compliers on different margins, can sign or eliminate the asymptotic bias without adopting a more restrictive model of judge behavior. Such restrictions may be motivated by specific institutional details, theory, or prior research.

In the second part of the paper, we turn to our main empirical contributions: estimating the impacts of noncarceral conviction and incarceration on future criminal justice involvement. We use 2SLS with the conviction propensity of judges as an instrument for conviction, while controlling for their incarceration propensity.[4] Analogously, we use judges' incarceration

---

[3]Here, we follow the literature in referring to an estimand as "causal" if it is a non-negatively weighted average of local average treatment effects (LATEs).

[4]This approach mirrors a common strategy used to study the impact of incarceration on recidivism. See

propensity as an instrument for incarceration and control for their dismissal propensity. Under the assumptions described in the first part of our discussion, our estimates imply that noncarceral conviction relative to dismissal leads to large and long-lasting increases in future criminal justice involvement, while incarceration relative to noncarceral conviction decreases recidivism in the first year, consistent with incapacitation.[5]

Our finding that noncarceral conviction increases recidivism (relative to dismissal) is consistent with both increased criminal behavior and an escalation in subsequent criminal justice responses. We examine how impacts differ by prior records, types of offenses, and measures of recidivism, but do not find evidence that supports one mechanism over the other. Both channels imply that a felony conviction can lead individuals to cycle back into the criminal justice system, leading to increased charges, convictions, and future incarceration.

To probe whether it is reasonable to interpret our 2SLS estimates as causal and margin-specific effects, we propose a test that evaluates whether the instruments are treatment specific. Since each model has different implications for treatment-specificity of the instruments, the test also lets us adjudicate between different models of judge decision-making. Our findings suggest that neither instrument is treatment specific. We can therefore empirically reject the ordered and sequential models of judge decision-making. Therefore, our 2SLS estimates could be biased. However, the magnitude of the bias can vary by context. We use our expression for the asymptotic bias, along with theory and empirical evidence, to argue that, in our setting, the bias is unlikely to overturn our qualitative conclusion regarding the effect of noncarceral conviction.

To assuage any remaining concerns about bias in the 2SLS estimates, we provide an alternative approach for identifying and estimating margin-specific treatment effects under the unordered choice model, which is not rejected by our test. We develop a novel approach that builds on Mountjoy (2022). This approach requires treatment-specific instruments, which we have argued judge stringencies generally are not. Following methods from the discrete choice literature, we impose additional structure on the judge's choice problem to construct treatment-specific instruments from judge stringencies. We then use these newly-constructed instruments to obtain estimates of margin-specific treatment effects. The results are similar to our 2SLS estimates, although they are smaller and less precise.

This research contributes to both applied and methodological literatures. First, our work is related to a small set of recent studies that explore the impact of criminal convictions. Mueller-Smith and Schnepel (2021) and Augustine et al. (2022) show that felony diversion causes large and sustained reductions in future criminal justice contact. Felony diversion helps avoid conviction, but can also affect recidivism through other channels. For instance, there may be enhanced deterrence, since rearrest leads to reinstated charges. In the context of misdemeanors, Agan, Doleac, and Harvey (2023) show that the decision to file charges

_____

Loeffler and Nagin (2022) and Doleac (2023) for recent reviews of this literature.

[5]We also examine the effects of incarceration using a regression discontinuity design based on sentencing guidelines, yielding conclusions that are consistent with our main findings.

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. In related work, Kamat, Norris, and Pecenco (2024) adopt a partial-identification approach and find that misdemeanor conviction increases the number of future charges, but they cannot rule out large effects of felony conviction in either direction. Additionally, there is a socio-legal literature providing theoretical arguments, as well as qualitative and descriptive evidence about the adverse effects of both felony and misdemeanor convictions (e.g., Chiricos et al., 2007; Natapoff, 2011; Phelps, 2017; Irankunda et al., 2020). We contribute to the existing literature by disentangling conviction from other aspects of the criminal justice process and by assessing the relative importance of felony conviction and incarceration in driving future criminal justice involvement within the same empirical setting.

Second, this paper contributes to the large body of work investigating the consequences of incarceration for recidivism. A recent review shows that post-conviction incarceration generally is not found to have long-term effects on recidivism, while pretrial detention increases recidivism after the incapacitation period (Loeffler and Nagin, 2022). Our study suggests one way to reconcile these findings: since pretrial detention increases the likelihood of conviction (Gupta, Hansman, and Frenchman, 2016; Leslie and Pope, 2017; Stevenson, 2018; Dobbie, Goldin, and Yang, 2018), adverse effects of pretrial detention may be operating through conviction rather than the experience of incarceration itself. Studies that identify the impacts of post-conviction incarceration, meanwhile, are often comparing incarceration to noncarceral conviction, with both the treatment and control groups being convicted.

We build on a methodological literature about the identification and estimation of treatment effects in the presence of multiple treatment alternatives. The prior and contemporaneous literature has outlined many of the challenges associated with multiple treatments (e.g., Heckman and Pinto, 2018; Heinesen et al., 2022; Bhuller and Sigstad, 2024; Kamat, Norris, and Pecenco, 2024). However, not all of the insights developed in the prior literature apply to the judge IV setting, given the special nature of judge stringency instruments as shares of cases that are allocated to particular court outcomes. Identification issues specific to judge IV in a multiple-treatment setting have received sustained consideration in two prior papers studying the impacts of incarceration. Mueller-Smith (2015) provides one of the first in-depth discussions of the challenges inherent in this design and proposes controlling for judge stringency along "non-focal" dimensions (such as fine amount or probation length). Arteaga (2021) discusses multiple-treatment identification issues and shows how to identify causal effects along the incarceration vs noncarceral conviction margin within a sequential model, which is a special case of our framework.

Our paper contributes to the methodological literature in several ways. First, we lay out identifying assumptions sufficient for 2SLS judge IV to yield a causal and margin-specific estimand when there are multiple treatments. In contemporaneous work, Bhuller and Sigstad (2024) present an alternative set of identifying conditions for 2SLS with multiple treatments. Their regression model is different: it instruments for all treatments simultaneously, and

thus requires stronger functional form assumptions than our approach. The monotonicity conditions they propose are weaker than ours, but ours have straightforward and tractable relationships with economic models of judge behavior. One of our contributions is to show how our econometric assumptions relate to three commonly used discrete choice models, which helps illuminate the econometric implications associated with different ways of modeling the court system. We also derive an expression for asymptotic bias under a weaker set of monotonicity assumptions that all of the choice models we consider satisfy. We suggest an empirical test for instrument treatment-specificity, and we demonstrate how to reason about the sign and magnitude of the bias term if the assumption is rejected. Finally, we show how to construct treatment-specific instruments from judge stringency instruments under an alternative set of assumptions, thus allowing the researcher to apply the identification approach presented in Mountjoy (2022) or other approaches that require such instruments (e.g., Lee and Salanié, 2018).

Lastly, our paper is related to a broad body of applied work that uses judge instruments. We offer a practical guide for research designs using such instruments when judges choose between more than two options.[6] Researchers can use their institutional knowledge to reason about which choice model fits best and apply the tests that we suggest to see if the data is consistent with their model. Our paper suggests that if both institutional expertise and the tests support an ordered model, 2SLS is a good choice. If either institutional knowledge or the empirical test reject the ordered model, then 2SLS estimands may have an additional bias term. In that case, theory and empirical results from prior literature can help the researcher reason about the sign and magnitude of the bias, as we demonstrate in our setting. Lastly, our alternative approach to identification can be used if institutional knowledge and empirical tests support an unordered model, and if the additional assumptions for constructing treatment-specific instruments are met. It can also be used as a robustness check to IV specifications.

Our paper proceeds as follows. Section II describes the institutional setting and our data. Section III extends the random judge design to multiple treatments and presents a set of sufficient conditions for 2SLS to recover causal and margin-specific treatment effects. We show how the treatment-specific instruments assumption rules out some commonly used models of discrete choice, and then derive an expression for the asymptotic bias if this assumption fails. Section IV presents our empirical results based on 2SLS estimates and introduces an empirical test for treatment-specific instruments. Section V describes an alternative approach to identification and estimation, as well as corresponding empirical results.

---

[6]Judge stringency instruments have been used in the criminal justice setting, but also in other settings, such as foster care (Doyle, 2008; Gross and Baron, 2022), disability claims (Maestas, Mullen, and Strand, 2013), bankruptcy (Dobbie and Song, 2015), eviction (Collinson et al., 2024), or patent decisions (Sampat and Williams, 2019). In many settings, decision-makers have multiple alternatives: pretrial detention, electronic monitoring, or release (Rivera, 2023); opioid prescription, other pain medication, no prescription; foreclosure, loan modification, no court action.

II   INSTITUTIONAL DETAILS AND DATA

II.A   Felony case processing in Virginia

This section describes felony criminal case processing in Virginia, focusing on adjudication within the Circuit Court, which is the primary data source for this paper.

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. At this hearing, the prosecutor must convince the judge that there is probable cause that the defendant committed a felony. This hearing 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 of citizens conducts an additional review of the evidence. 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.[7] Our analyses include only 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 is assigned to a judge. The exact assignment procedure varies by jurisdiction, as we learned during phone interviews with court clerks. 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 whether the case number is odd or even. Appendix E shows that our results are robust to which case assignment mechanisms 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 in front of 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. In a bench trial, the judge decides whether to convict and, if so, what sentence to give. Judges also exert substantial indirect influence on adjudication and sentencing. For instance, judges decide what evidence is admissible, what charges can proceed, what must be struck from the record,

---

[7]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.

and what instructions the jury receives. Many of these decisions are made prior to trial. Since judges 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 latitude to reject any negotiated plea deal and to deviate from the sentence guidelines if they provide a written explanation. For all these reasons, judges can influence both the likelihood of conviction and incarceration.[8]

Virginia's criminal justice system compared to other states. Appendix A compares aggregate statistics for 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 a similar racial and ethnic composition of its incarcerated population. However, it has lower than average parole rates because Virginia adopted "truth in sentencing" for felony convictions starting in 1995, which requires people with felony convictions to serve at least 85% of their prison term. As a result, the initial carceral sentence is much more closely linked to time spent incarcerated than in other places.

II.B How noncarceral conviction and incarceration may affect recidivism

Noncarceral conviction. Receiving a felony conviction instead of a dismissal could increase or decrease recidivism through several 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, increasing the probability of apprehension. It could also raise future sentences conditional on conviction, since prior convictions are used to determine recommended sentences. Both of these channels suggest that noncarceral conviction increases the expected punishment for future offenses, thereby raising the costs of crime and potentially dampening recidivism (Drago, Galbiati, and Vertova, 2009; Philippe, 2024).

Alternatively, felony convictions may increase recidivism due to the stigma and desta-bilization associated with such records.[9] For instance, employers or landlords conducting background checks may be dissuaded from hiring or renting to someone with a felony con-viction, 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

---

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

[9]Our paper focuses on felony charges. While misdemeanor charges are more common (Mayson and Stevenson, 2020), they generally carry fewer legal and extra-legal consequences (Agan et al., 2024).

rates (see e.g. Pager, 2003; Holzer, Raphael, and Stoll, 2007; Agan and Starr, 2018; Craigie, 2020).[10]

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, 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. For example, a prior conviction could influence the likelihood that someone will be detained pretrial, or the prosecutor's willingness to offer diversion or bargain the charges down to a misdemeanor. Criminal justice actors have access to a defendant's full criminal record at nearly all stages of decision-making. While even prior arrests that were not sustained influence decisions (Kohler-Hausmann, 2018), convictions are generally considered more serious indicators of prior criminal behavior.

Incarceration. Incarceration could affect recidivism through several channels. It could reduce future criminal justice contact through incapacitation (Avi-Itzhak and Shinnar, 1973).[11] Incarceration could also decrease recidivism through specific deterrence (Zimring, Hawkins, and Vorenberg, 1973; Drago, Galbiati, and Vertova, 2009). 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 in the 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, Hjalmarsson, and Pozen, 2009; Stevenson, 2017).

II.C 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 Appendix B. This subsection also presents summary statistics.

Data. Our primary data source for the judge IV analysis in Section IV comes from Virginia's

---

[10]Both arrests and convictions are visible on background checks and both may influence employers' and landlords' decisions. However, since convictions have met a higher burden of proof, convictions are likely considered more serious than arrests that do not lead to conviction, in particular by employers (Agan et al., 2024). Note also that employment background checks submitted to the Virginia criminal records database do not show arrests that did not lead to a conviction (see VA Code §19.2-389).

[11]This doesn't mean that incarceration prevents crime, since crime is common 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.

Circuit Courts (2021). The data were 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 (hearing 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 (2021), which covers everyone convicted of a felony in Virginia during the period 1996–2020.

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 with substantial missing data or 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 (i.e., noncarceral conviction). Lastly, we say that a person was "dismissed" if all of their charges led to a dismissal 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.[12] Our main recidivism measure does not include probation revocations unless these are accompanied by a new felony charge for a new crime. We calculate recidivism in the first year, years two to four, years five to seven, and the first seven years after a person's initial conviction. We also consider two alternative measures of recidivism: a new conviction resulting from felony Circuit Court charges, or a new carceral sentence resulting from felony Circuit Court charges.

Summary statistics. Table I provides summary statistics for those dismissed, with a non-carceral conviction, or incarcerated, respectively. In our sample, 55% of cases ended with incarceration, about 30% ended with noncarceral conviction, and 15% ended with dismissal. The three groups are similar in terms of zip code-level poverty but differ demographically. Cases ending in a noncarceral conviction are more likely to have female and non-Black defendants. Cases ending in incarceration are more likely to have defendants with prior felony convictions (22%) compared to the noncarceral conviction and dismissed samples (10% and 14%, respectively). Drug charges are the most common charges for all groups, followed by larceny, assault, and fraud. Appendix Figure E.1 presents disposition types for four common offenses: drugs, larceny, assault, and fraud. While there is variation in the breakdown, all three disposition types exist within each offense type.

---

[12]Crimes committed during incarceration are usually addressed with internal sanctions and are unlikely to result in new felony charges. Hence, our main recidivism measures are likely to overlook crimes committed behind bars.

III EXTENDING BINARY-TREATMENT JUDGE IV TO MULTIPLE TREATMENTS

In this section, we extend the "random judge" framework from the binary-treatment setting to the setting with three possible court outcomes. We outline assumptions under which widely-used 2SLS regressions recover margin-specific treatment effects, provide intuition for their restrictiveness by relating them to models of judge decision-making, and derive an expression that can be used to reason about the likely sign and direction of bias when some of the assumptions are not met.

III.A Notation and common regression specifications

We consider a setting where cases can end in one of three mutually exclusive and collectively exhaustive alternatives: dismissal (d), noncarceral conviction (c), or incarceration (i). We denote treatment by $T \in \{d, c, i\}$. To simplify the discussion below, we further define $T_k \equiv \mathbb{1}\{T = k\}$ as an indicator for the outcome of the case being $k \in \{d, c, i\}$ and $T_{\setminus d} \equiv \mathbb{1}\{T \in \{c, i\}\}$ as an indicator that is equal to one if an individual is convicted or incarcerated (i.e., their case is not dismissed). Finally, we let Y be a measure of recidivism.

Both T_c and T_i are likely to be affected by unobserved factors that also influence recidivism, such as the strength of the evidence or the details of the offense or criminal record. Therefore, in a regression of Y on these court outcomes, there is concern about selection bias. To deal with selection, 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 $Z_i \equiv E[T_i \mid J]$ and let $z_i^j \equiv E[T_i \mid J = j]$, where $j \in \{1, ..., \mathcal{J}\}$ indexes the judges. Similarly define Z_k and z_k^j for $k \in \{c, d\}$.

Using the notation above and abstracting away from covariates, the following regression model is commonly used to study the impacts of incarceration (see, e.g., Mueller-Smith, 2015; Bhuller et al., 2020; Arteaga, 2021; Norris, Pecenco, and Weaver, 2021):

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

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

Estimated using two-stage least squares (2SLS), researchers instrument incarceration with the assigned judge's incarceration stringency, and controls for dismissal stringency Z_d to prevent exclusion violations stemming from the judge's likelihood of conviction.[13]

---

[13]Another common specification uses a second stage with the two endogenous treatments, T_c and T_i, instrumented with both stringencies, Z_c and Z_i. Under assumptions A1–A4, defined below, this specification produces the same 2SLS estimand as (1)–(2) (see Appendix C.6). Alternatively, researchers may consider including a single binary treatment indicator (e.g., T_i) and instrument with judge stringency in that dimension, omitting controls for other dimensions of sentencing. Under the standard Imbens and Angrist (1994) LATE assumptions, this approach does not recover a well-defined causal effect when there are multiple treatments and the stringencies are correlated, which is likely given that $Z_i = 1 - (Z_c + Z_d)$ (see Appendix C.7).

Analogously, one approach to learning about the impacts of a noncarceral conviction is to use the following specification, which instruments for conviction but controls for incarceration stringency:

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

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

In the next subsection, we discuss conditions under which δ_1 has a causal and margin-specific interpretation—i.e., when it can be interpreted as the impact of noncarceral conviction relative to dismissal for some well-defined subgroup of the population.[14]

III.B **Judge IV assumptions in 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 $Y(t, z_i, z_c)$, $t \in \{d, c, i\}$. We further define $T_k(z_c, z_i) \equiv \mathbb{1}\{T(z_c, z_i) = k\}$, for $k \in \{d, c, i\}$. Using this notation, we can state the standard IV assumptions of exclusion, random assignment, and relevance for 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: $\gamma_1 \neq 0$ in equation (3).

We additionally make the following functional form assumption, following Blandhol et al. (2022):

A4. Rich covariates: The linear projection of Z_c on Z_i is equal to $E[Z_c \mid Z_i]$.

Equations (3)-(4) instrument for conviction using Z_c while controlling for Z_i rather than instrumenting for conviction and incarceration jointly in the same 2SLS regression. An advantage of this approach is that concerns about the validity of Assumption A4 can be alleviated in a straightforward way by controlling for Z_i more flexibly (see Appendix C.5).

Throughout the paper, unless specified otherwise, we assume A1–A4 are satisfied. A1–A3 represent straightforward analogs to the standard Imbens and Angrist (1994) assumptions, and A4 implements a functional form assumption related to using Z_i as a control. Extending the monotonicity assumption from the binary- to the multiple-treatment setting is less straightforward. In other applications, researchers have assumed that instruments induce

---

[14]Our discussion in the reminader of this section is organized around the interpretation of δ_1 in specification (3)–(4), but an analogous argument holds for β_1 in specification (1)–(2).

compliers to take up a specific treatment, without inducing anyone to switch into other "non-focal" treatments, i.e. that the instruments are treatment-specific. For example, Kline and Walters (2016) study the impact of enrolling in Head Start in a setting with two outside options, using randomly-assigned offers of enrollment as an instrument. The Head Start offer is assumed to not induce switches between the outside options. Similarly, Kirkeboen, Leuven, and Mogstad (2016) study the returns to college majors and use offers of admission to specific majors as instruments. Their irrelevance condition states that access to a major does not induce switches between other choices (e.g., increased access to an economics major won't induce students to switch between history and mathematics). In a similar vein, Mountjoy (2022) assumes that reducing the distance to a two-year college (while holding distance to four-year college fixed) lowers its relative costs, but does not induce switches between four-year college and not enrolling.

The Unordered Partial Monotonicity (UPM) assumption in Mountjoy (2022) formalizes the treatment-specific instruments assumption. In our notation, this assumption can be stated as:

A5. Unordered Partial Monotonicity (UPM($Z_c \mid Z_i$)):

For all z_c, z_c', z_i with $z_c' > z_c$ and holding z_i fixed:

   i. $T_c(z_c', z_i) \geq T_c(z_c, z_i)$,
   ii. $T_i(z_c', z_i) \leq T_i(z_c, z_i)$,
   iii. $T_d(z_c', z_i) \leq T_d(z_c, z_i)$.

Treatment specificity of an instrument for conviction, as formalized by UPM, imposes three restrictions on substitution patterns when Z_c increases and Z_i is held fixed. First, it guarantees that individuals only move into (and not out of) noncarceral conviction. Second, it guarantees that individuals only (weakly) move in one direction across any margin. Third, it rules out flows between dismissal and incarceration.[15] The UPM assumption thus incorporates a property similar to the "no defiers" assumption in the binary setting (Imbens and Angrist, 1994), but additionally rules out switches between incarceration and dismissal.

When using judge stringencies as instruments, the UPM assumption restricts substitution patterns more than in the three studies discussed above. In those examples, the instruments reduce costs or increase access to specific choices. In contrast, judge stringency instruments are the judge-specific probabilities of a case ending with a particular outcome. The stringency instruments will add up to one ($z_d^j + z_c^j + z_i^j = 1$) since our case outcomes are mutually exclusive. As such, judge stringency instruments vary the probabilities of taking up particular treatments. If we condition on the judge stringency for one particular treatment, we fix its net probability of take-up.

---

[15]Note that UPM can hold when varying one instrument and holding the other fixed, but not hold when switching the roles of the instruments. We therefore use the notation UPM($Z_c \mid Z_i$) for the definition above and UPM($Z_i \mid Z_d$) when incarceration is the focal treatment.

This feature of judge instruments is important for understanding judge IV with multiple treatments. If we increase conviction stringency Z_c while holding Z_i fixed, we increase the net probability of conviction while holding the net probability of incarceration constant. Thus, if increasing Z_c results in an $i \to c$ shift, there must also be a compensating same-sized $d \to i$ shift to keep the net probability of incarceration constant. However, UPM($Z_c \mid Z_i$) rules out flows from dismissal to incarceration. Since the net probability of incarceration Z_i is held fixed, there can be no $i \to c$ flows. Therefore, UPM implies that judge stringency instruments are not only *treatment-specific*, as in the examples described above, but also *margin-specific*: they induce complier flows across only one margin, e.g., dismissal to noncarceral conviction. In the multiple-treatment judge IV setting, UPM therefore ensures that we recover margin-specific treatment effects, but it is also a less plausible assumption than in many other multiple-treatment IV settings. In Section *III.C* we illustrate the restrictiveness of the UPM assumption by showing that it rules out certain reasonable models of judge decision-making.

Given that UPM may be a particularly strong assumption with judge stringency instruments, we next introduce a weaker monotonicity assumption, which we call conditional pairwise monotonicity (CPM).[16]

A6. Conditional Pairwise Monotonicity (CPM($Z_c \mid Z_i$)):

For case outcomes c, i, and d, for all z_c, z_c', z_i with $z_c' > z_c$ and holding z_i fixed:

  i. $T_c(z_c', z_i) \geq T_c(z_c, z_i)$ for all individuals

 ii. if $T_i(z_c', z_i) = T_d(z_c, z_i) = 1$ for any individual, then $T_i(z_c, z_i) = 1$ implies $T_d(z_c', z_i) = 0$ for all individuals

iii. if $T_d(z_c', z_i) = T_i(z_c, z_i) = 1$ for any individual, then $T_d(z_c, z_i) = 1$ implies $T_i(z_c', z_i) = 0$ for all individuals.

CPM imposes two of the three restrictions imposed by UPM. It guarantees that, in response to increasing Z_c while holding Z_i fixed, individuals only move into (and not out of) $T = c$ and that individuals only (weakly) move in one direction across any margin.[17] CPM does not rule out flows across margins that are not adjacent to noncarceral conviction. For example, an increase in Z_c holding Z_i constant can induce $d \to c$ and $i \to c$ flows, but also $d \to i$ flows. Throughout this paper, we assume CPM holds. Next, we discuss the implications for 2SLS estimands when CPM holds but UPM does not.[18]

---

[16]Another way to relax the UPM assumption would be to extend the concept of average monotonicity (Frandsen, Lefgren, and Leslie, 2023) to the multiple-treatment setting. We present a definition of "average UPM" in Appendix C.4. Bhuller and Sigstad (2024) provide a more general way to extend average monotonicity with an arbitrary number of treatments. They provide conditions that are both sufficient and necessary for a 2SLS estimand to have "proper weights."

[17]Note that conditions (ii) and (iii) in A6 can be replaced with $T_d(z_c', z_i) \leq T_d(z_c, z_i)$ within our setting with stringency instruments, which makes CPM equivalent to (i) and (iii) from the UPM definition.

[18]While CPM is weaker than UPM, it still imposes restrictions on judge behavior that may not hold: it

III.C   Connecting assumptions to models of judge decision-making

In this subsection, we provide economic intuition for the assumptions in the previous subsection, by examining how they restrict models of judge decision-making. We consider three index-crossing models of judge decision-making based on canonical models of multinomial discrete choice—an ordered choice model, a sequential choice model, and an unordered choice model—and discuss how they relate to the legal and institutional practices of criminal proceedings.[19] All three models satisfy the CPM assumption. Only the ordered choice model satisfies the UPM assumption for both instruments. The sequential model illustrates that UPM may be satisfied for one of the instruments but not the other.

III.C.1   Ordered choice

First, we consider a straightforward extension to a trinary model from the binary threshold-crossing model. This extension is 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, noncarceral conviction, and incarceration:

$$
\begin{aligned}
T_d &= \mathbb{1}\{W < \pi_c(Z_d)\} \\
T_c &= \mathbb{1}\{\pi_c(Z_d) \leq W < \pi_i(Z_i)\} \\
T_i &= \mathbb{1}\{W \geq \pi_i(Z_i)\}.
\end{aligned}
\tag{5}
$$

Panel (a) in Figure I visualizes, for two different judges, the regions of W under which each judge dismisses, convicts, and incarcerates. In this example, judge 1 has higher thresholds for both noncarceral conviction and for incarceration than judge 2.

In an ordered choice model, we can estimate margin-specific treatment effects for both the conviction-dismissal margin and the incarceration-conviction margin. Consider panel (b) of Figure I, where both judges have the same incarceration threshold, but judge 2 has a lower noncarceral conviction threshold, meaning that they convict more and dismiss less than judge 1. Fixing Z_i and increasing Z_c will hold $\pi_i(Z_i)$ fixed and decrease $\pi_c(Z_d)$. The only people who will switch treatment status are those who move from dismissal to conviction. When 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 ($d \to c$). Similarly, we can learn about the effect of incarceration vs noncarceral conviction using variation in Z_i and fixing Z_d. Thus, this choice model satisfies the unordered partial monotonicity assumption for both

---

rules out defiers by requiring the instrument moves everyone in the same direction across a margin (see e.g. Chaisemartin, 2017; Chan, Gentzkow, and Yu, 2022; Frandsen, Lefgren, and Leslie, 2023; Sigstad, 2023, for more elaborate discussions). Here, we focus on the novel issues that arise with judge stringency instruments and multiple treatments.

[19]Throughout this subsection we use "models of judge decision-making" as a shorthand; in practice, court outcomes reflect a combination of decisions by multiple actors.

margins (i.e., UPM($Z_c \mid Z_i$) and UPM($Z_i \mid Z_d$) hold).

This model would be appropriate if all judges considered a single dimension of unobserved heterogeneity in their decision, and they agreed on how cases are ranked according to this dimension. The only way judges can differ in their decision making is by setting different thresholds for assigning cases to each outcome. In practice, however, judges could take into account more than one measure of unobserved heterogeneity. In the remainder of this section, we consider models that allow for multiple dimensions of unobserved differences between defendants.

III.C.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 model 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. It allows judges to consider different, though potentially correlated, unobserved factors in each decision. For example, conviction decisions may depend on the strength of the evidence, while incarceration decisions may depend on other aspects, such as the propensity to re-offend or severity of the crime.

We can write this as a threshold-crossing model:

$$
\begin{aligned}
T_d &= \mathbb{1}\{U_c < \pi_c(Z_d)\} \\
T_c &= \mathbb{1}\{U_c \geq \pi_c(Z_d), U_i < \pi_i(Z_i, Z_d)\} \\
T_i &= \mathbb{1}\{U_c \geq \pi_c(Z_d), U_i \geq \pi_i(Z_i, Z_d)\}.
\end{aligned}
\tag{6}
$$

In this model, the first choice is between $T \in \backslash d$ (not dismissed) and $T = d$ and depends on the value of case-specific unobservable U_c relative to judge-specific threshold π_c. For cases that switch from dismissed to "not dismissed," there is then a second choice: noncarceral conviction or incarceration. This choice depends on the value of case-specific unobservable U_i, which can be correlated with U_c, relative to judge-specific π_i.[20] This model is consistent with different factors being relevant at each stage of the decision. For example, evidence might be more relevant to conviction while the criminal record might be more relevant to sentencing. 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 sequential model and A1-A4, it is possible to use 2SLS and the stringency instruments to recover margin-specific treatment effects for the incarceration-conviction margin, but not for the conviction-dismissal margin or the dismissal-no dismissal margin. Figure II illustrates this point. Panel (a) visualizes the decision regions of one judge, which are based on U_c and U_i. Panel (b) compares two judges who have the same probability

---

[20]See Heckman, Humphries, and Veramendi (2016) for details on identifying 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.

of dismissal, but where the second judge has a higher probability of incarceration. Here, variation in Z_i holding Z_d fixed only induces $c \to i$ changes in court outcomes 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 resulting in incarceration. In this figure, Z_i is represented by the fraction of people in the top right section. For two judges to have the same incarceration shares but different dismissal probabilities, both π_i and π_c must differ across these judges. Comparisons across these two judges induces three sets of compliers: $d \to c$ compliers, $i \to c$ compliers, and $d \to i$ compliers. This example satisfies CPM since there is only a one-way flow across any margin, and no flows out of treatment. However, the flow from $T = d$ to $T = i$ implies that the instrument is not treatment-specific, and UPM($Z_c \mid Z_i$) is not satisfied.

Although the sequential model captures the two-step nature of the criminal proceeding, it may not be a good model if the outcome of the case is determined by a *joint consideration* of the two dimensions, as may be the case when plea bargaining occurs. We thus also consider a multinomial choice model, which has two dimensions of unobserved heterogeneity, but allows for both unobservables to affect both conviction and incarceration.

III.C.3   Unordered multinomial choice

We now consider an unordered multinomial choice model, where outcomes can be thought of as being determined by judges maximizing over their "returns":[21]

$$
\begin{aligned}
R_c &\equiv V_c - \pi_c(Z_c, Z_i) \\
R_i &\equiv V_i - \pi_i(Z_c, Z_i) \\
R_d &\equiv 0.
\end{aligned}
\tag{7}
$$

The outcome of the case depends on the judge's threshold for noncarceral conviction ($\pi_c(Z_c, Z_i)$), the judge's threshold for incarceration ($\pi_i(Z_c, Z_i)$), and two unobserved case-specific characteristics (V_c and V_i). Thus, this model assumes that case outcomes are determined by a joint consideration across the two unobserved dimensions, which may better capture the intertwined decisions that are common in Virginia and other US jurisdictions due to plea bargaining. In a plea deal, a defendant typically agrees to plead guilty in exchange for a lower sentence, making conviction and sentencing determinations closely connected; unobserved determinants of the sentencing decision may affect the decision to plead guilty.

Using the unordered multinomial choice model above, we can write the three treatment

---

[21]See, e.g., Heckman, Urzua, and Vytlacil (2006) for a discussion of treatment effects in an unordered multinomial choice model and Mountjoy (2022) for an application in the context of college choice.

indicators as:

$$
\begin{aligned}
T_d &= \mathbb{1}\{V_c < \pi_c(Z_c, Z_i), V_i < \pi_i(Z_c, Z_i)\} \\
T_c &= \mathbb{1}\{V_c \geq \pi_c(Z_c, Z_i), V_c - V_i \geq \pi_c(Z_c, Z_i) - \pi_i(Z_c, Z_i)\} \\
T_i &= \mathbb{1}\{V_i \geq \pi_i(Z_c, Z_i), V_i - V_c \geq \pi_i(Z_c, Z_i) - \pi_c(Z_c, Z_i)\}.
\end{aligned}
\tag{8}
$$

The propensity of a judge to convict depends on both π_i and π_c, neither of which is directly observed. Panel (a) of Figure III visualizes the court outcomes and how they depend on judge thresholds and the two unobservables.

Under this model, the instruments are not treatment-specific. Consider panel (b) of Figure III, which shows how treatment assignment changes when holding Z_i fixed and increasing Z_c. In this case, individuals shift from incarcerated to convicted and from dismissed to convicted but, to hold the probability of incarceration (Z_i) constant, individuals also need to 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. Similar conclusions are drawn when holding Z_c (or Z_d) fixed and varying Z_i. Hence, under this model, 2SLS with stringency instruments does not recover margin-specific or treatment-specific treatment effects without further assumptions.

These observations illustrate how judge stringency instruments with multiple treatments differ from those in Mountjoy (2022), which also considers an unordered choice model. The difference stems from the fact that stringency instruments are generally not treatment-specific, while the distance instruments in Mountjoy (2022) plausibly are. The judge stringency for conviction does not correspond to π_c; it corresponds to the fraction of court cases in the conviction section of the graph. By contrast, the distance instrument in Mountjoy (2022) directly shifts π_c, holding π_i constant. In our setting, this 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 III. Given that we do not observe π_c or π_i, we 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 treatment.

III.D Asymptotic bias under different monotonocity assumptions

The prior subsection showed how UPM rules out some reasonable models of judge behavior, while the weaker CPM condition is not sufficient for 2SLS to recover margin-specific or treatment-specific effects. Here we derive the Wald estimand under CPM, which is satisfied by all three models. As in the prior section, we will consider the impacts of conviction vs dismissal and study the case where Z_c takes two values and Z_i is fixed. Analogous results for the incarceration-conviction margin can be obtained by rearranging subscripts.

Consider increasing conviction stringency from z_c to z_c' while holding incarceration stringency fixed at z_i. Let $\omega_{i \to c}$ represent the proportion of cases switching from incarceration to conviction in response to the instrument shift. Similarly, allow $\omega_{d \to c}$ and $\omega_{c \to i}$ to represent

the proportions of cases responding by switching across the other margins. Next, let $\Delta_{i \to c}^{Y_c - Y_i}$ represent the local average of $T = c$ vs $T = i$ treatment effect for $i \to c$ compliers when the instrument shifts from z_c to z_c', holding Z_i fixed. More generally, $\Delta_{k \to l}^{Y_m - Y_n}$ denotes the treatment effect of $T = m$ vs $T = n$ for $k \to l$ compliers.[22]

Proposition 1. *Under A1–A4 and CPM, 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*

$$\frac{E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} =$$

$$\underbrace{\frac{\omega_{d \to c} \Delta_{d \to c}^{Y_c - Y_d} + \omega_{i \to c} \Delta_{d \to i}^{Y_c - Y_d}}{\omega_{d \to c} + \omega_{i \to c}}}_{\text{Positively-weighted avg. of } Y_c - Y_d \text{ treatment effects}} + \underbrace{\frac{\omega_{i \to c}}{\omega_{d \to c} + \omega_{i \to c}} \left[\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c} \right]}_{\text{Bias term}}. \quad (9)$$

Proof: See Appendix C.1.

Proposition 1 states that the Wald estimand can be decomposed into two terms. The first term is a weighted average of two LATEs for noncarceral conviction vs dismissal, corresponding to two different groups of compliers. The second term represents asymptotic bias relative to this weighted average. The bias term is the difference between the LATE for incarceration vs conviction for two equally-sized groups of compliers, weighted by the share of compliers moving from incarceration to noncarceral conviction. A direct consequence of Proposition 1 is that, when we replace the CPM assumption with the UPM assumption, the bias term in equation (9) is eliminated.

Corollary 1. *Under A1–A4 and UPM, 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*

$$\frac{E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} = E[Y(T(z_c', z_i)) - Y(T(z_c, z_i)) \mid T_c(z_c', z_i) = T_d(z_c, z_i) = 1]$$

$$= \Delta_{d \to c}^{Y_c - Y_d}. \quad (10)$$

This corollary stems from the fact that the bias term is zero if $\omega_{i \to c}$ equals zero, i.e. no compliers shift from incarceration to conviction. As discussed previously, UPM combined with judge stringency instruments ensures that this condition is met. Thus, under UPM, the Wald estimand will be $\Delta_{d \to c}^{Y_c - Y_d}$, which is the LATE for noncarceral conviction vs dismissal for those shifted across that margin by the instrument.

Proposition 1 and Corollary 1 allow us to reason about conditions under which asymptotic bias will be quantitatively important for our 2SLS estimands. Under A1–A4 and UPM, the 2SLS specification in equations (3)–(4) yields a positively-weighted sum of unbiased

---

[22]For simplicity, we suppress notation indicating instrument values; for example, we write $\omega_{d \to c}$ rather than $\omega_{d \to c}(z_c', z_c \mid z_i)$ and $\Delta_{j \to k}^{Y_m - Y_n}$ rather than $\Delta(z_c', z_c \mid z_i)_{j \to k}^{Y_m - Y_n}$.

Wald estimands.[23] If CPM holds but UPM does not, then the 2SLS estimands will represent a positively-weighted sum of the *biased* Wald estimands from equation (9) unless we impose additional assumptions. One possibility is to restrict treatment effect heterogeneity.

Treatment effect homogeneity assumptions under which the bias term is zero. The bias term will be zero if $\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c} = 0$. Thus, if the average treatment effects of incarceration vs conviction are the same for the $d \to i$ compliers and $i \to c$ compliers, the bias will be zero. For this result, we do not need the stronger assumption that treatment effects are homogeneous across all cases. Nor do we need to assume treatment effect homogeneity across the conviction-dismissal margin.[24] A special case occurs when the impact of incarceration vs conviction is zero for these two groups. This case is of specific interest in our context, because prior studies find null effects across this margin after the incapacitation period (see, e.g., Loeffler and Nagin, 2022; Garin et al., 2024). We return to this point in Section *IV.E.2* .

Reasoning about sign and magnitude of the bias. Equation (9) also allows us to reason about the likely sign and magnitude of the bias when we are unwilling to make the homogeneity assumptions discussed above. We know that the bias term is less than and proportional to $\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c}$, i.e., the difference in the impact of incarceration (relative to noncarceral conviction) between $d \to i$ compliers and $i \to c$ compliers. Thus the sign and the magnitude of the bias depend on the differential impact across these two groups. We illustrate how it is possible to reason about this differential impact in our context of criminal court cases in Section *IV.E*.

IV CONVICTION, INCARCERATION, AND RECIDIVISM: 2SLS ESTIMATES

IV.A Regression specifications for estimation

Using leave-one-out estimates of judge stringency as our instruments, we consider the following 2SLS regression model, which is common in the literature (stated here for noncarceral conviction; the specification for incarceration is analogous):

$$T_c = \delta_0 + \delta_1 Z_c + \delta_2 Z_i + \delta_3' X + \epsilon \tag{11}$$

$$Y = \gamma_0 + \gamma_1 T_c + \gamma_2 Z_i + \gamma_3' X + \nu, \tag{12}$$

---

[23]Note that assumptions A1–A5 imply the assumptions needed in Blandhol et al. (2022) for 2SLS to recover causal estimands. In particular, A5 implies their "Ordered strong monotonicity" (OSM). See Appendix C.3 for details, and see Appendix C.5 for how to interpret the 2SLS estimand when additional covariates are included in the 2SLS regression.

[24]Also note that homogeneous treatment effects rule out selection on gains, but still allow for selection on level (e.g., individuals more prone to recidivism can be more likely to be incarcerated).

where Y is one of the measures of recidivism described in Section $II.C$. The vector X includes court-by-year, court-by-month, 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 main measure of judge stringency, we use the three-year leave-one-out conviction and incarceration rates for the judge handling the case.[25] We run these 2SLS regressions on the sample described in Section $II.C$.[26]

In Appendix D, we discuss how assumptions A1–A3 are supported by features of the institutional environment and provide results from empirical tests of these assumptions. For both the conviction and incarceration regressions, we have a strong first stage with F-statistics of 165 and 288 (Table II), suggesting that relevance holds in our setting. Panels A and B of Figure IV plot the variation in residualized judge conviction and incarceration stringency, showing that there is substantial variation in each. Panel C of Figure IV provides a scatter plot of residualized conviction and incarceration stringency and shows that there is also substantial variation in Z_c conditional on Z_i, and vice versa. For balance, Table III shows that, while case characteristics are strong predictors of conviction and incarceration, they do not predict judge stringencies. For the few covariates with statistically significant loadings, the predicted difference in stringency tends to be very small (0.016 to 0.036 standard deviations of the residualized stringency measure, see Appendix Table D.1). In addition, Appendix Tables D.2 and D.3 show that our main results are broadly similar when systematically dropping certain case types, such as assault. For the exclusion restriction, we discuss potential violations and provide tests suggesting that these would not have qualitative impacts on our results. For instance, we show in Figures E.3–E.6 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 Table D.5 reporting split-sample monotonicity tests and finding the same sign for the first stage across various splits of the data. We postpone the discussion and implementation of an additional test of the UPM assumption to Section $IV.E$.

IV.B **Noncarceral conviction**

Table IV presents 2SLS estimates of the model in equations (11)–(12). 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, and cumulatively for the

---

[25]We choose a three-year window to ensure that the stringency measures are computed based on an adequate 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 three-year period.

[26]As discussed in Section III, under A1–A5, these regression estimates can be interpreted as causal and margin-specific. See Appendix C.5 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.

first 7 years. For each of these outcomes, we present OLS and 2SLS regressions.[27]

As discussed in Section II.B, noncarceral conviction (instead of a dismissal) could increase or decrease recidivism through multiple 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 noncarceral conviction increases future criminal justice contact relative to dismissal. The estimates for future charges within the first year after conviction are large: around 10.5 percentage points (95% CI, 0.02 to 0.20), which is a 67% increase relative to the control complier mean. The impact on cumulative recidivism 1–7 years later is also statistically significant, with an estimate of 24 percentage points (95% CI, 0.05 to 0.43), a 48% increase relative to the control complier mean. The effects for years 1–7 are approximately twice as large as the effects in year 1, with positive but statistically insignificant effects in years 2–4 and 5–7. The results are similar for the other measures of recidivism we consider.

These point estimates are similar in magnitude to estimates found in the related literature. For instance, Mueller-Smith and Schnepel (2021) find that diversion cuts reoffending rates in half, and Agan, Doleac, and Harvey (2023) find that nonprosecution reduces the likelihood of a new criminal complaint by 53%. Mueller-Smith, Pyle, and Walker (2023) find that adult conviction increases the total number of future felony charges by roughly 75%. While our point estimates could be considered fairly large, the confidence intervals leave room for a wide range of values, as is typical for judge IV research designs.

Our 2SLS estimates are similarly signed but substantially larger than the OLS estimates. However, the OLS estimates likely suffer from omitted variable bias. One important omitted variable is the strength of the evidence, which often consists primarily of witness testimony. Graef et al. (2023) 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—for example, poverty or social marginalization—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, OLS estimates will be downward biased.[28] Alternatively, IV compliers may be more impacted by conviction than the average defendant. In Appendix

---

[27]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.

[28]Witness cooperation is just one potential omitted variable among many that could bias the OLS estimates. For example, if people with skilled lawyers are both less likely to be convicted and to recidivate, OLS would be upward biased; conversely, if those with untreated substance issues are less likely to be convicted but more likely to recidivate, OLS would be downward biased.

Table E.2, we show that the racial composition of the complier group is similar to the overall sample, but that on average this group is less likely to be in court for violent offenses and is less likely to have a prior conviction. Our OLS estimates for noncarceral conviction are somewhat larger when reweighting with complier weights, while the estimates for incarceration do not notably change (see Appendix Table E.3).

We next consider several mechanism. Noncarceral conviction (relative to dismissal) could affect recidivism due to fines and probation conditions. However, a small but growing literature shows that court fines and fees do not affect recidivism (Pager et al., 2022; Finlay et al., 2024; Lieberman, Luh, and Mueller-Smith, 2023). Similarly, several large-scale RCTs have shown that probation and parole conditions do not affect recidivism (Doleac, 2023). We therefore focus on asking whether our results are coming from an increase in criminal behavior or an escalation in subsequent responses by the criminal justice system ("ratcheting up")—mechanisms we discussed in Section II.B. While we cannot answer this question definitively, we conduct two tests to provide suggestive evidence.

First, 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.4 and find similar point estimates across income-generating and non-income-generating crime; the confidence intervals are too large to draw a firm conclusion.[29] Second, 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. The logic here is that if a felony conviction increases the likelihood of a negative outcome at each discretionary stage, the negative impact of a conviction will accumulate. Downstream outcomes, like incarceration, will be impacted more than upstream outcomes, like the charging decision. Comparing the three measures of recidivism in Table IV, the point estimates are larger relative to the control complier means for outcomes with more discretionary decisions, providing suggestive evidence that the ratcheting up mechanism is present.

While we cannot conclusively say whether increased recidivism is driven primarily by increased criminal behavior or a ratcheting-up effect, both mechanisms imply that felony conviction contributes to the revolving door of criminal justice, increasing not just future charges and convictions, but also future incarceration.

IV.C Incarceration

Table V presents 2SLS estimates of the model analogous to those in equations (11)–(12), but instrumenting for incarceration with incarceration stringency and controlling for dismissal stringency.

When given a causal and margin-specific interpretation, our 2SLS estimates suggest a 10 percentage point reduction in future charges in the first year (95% CI, -0.15 to -0.04).

---

[29]Likewise, there are no consistent differential patterns for drug vs. non-drug crimes, as shown in Appendix Table E.5.

This reduction is likely due, at least partially, to incapacitation. We find no evidence that incarceration affects future criminal justice interactions beyond the first year. The 2–4 year and 5–7 year estimates are small and statistically insignificant. The cumulative estimate across all seven years implies a seven percentage point reduction in new felony charges (95% CI, −0.19 to 0.05). We can reject increases in recidivism larger than 2.6 percentage points at the .05 level. Results are similar for future convictions and future incarceration.

Our qualitative conclusions regarding incarceration effects are further strengthened by the fact that we find similar results using another research design within the same institutional setting. We leverage the fact that judges' sentencing decisions are influenced by sentence guidelines. The guidelines-recommended sentence is calculated using a scoring system in which various characteristics of the offense and criminal record are assigned points which are then summed to create the sentence guidelines score. Exploiting two different discontinuities in the sentence guidelines recommendations within a regression discontinuity design framework, we estimate the effects of incarceration on the intensive margin (sentence length) and on the extensive margin (short jail sentences vs probation). As when exploiting quasi-random assignment of cases to judges, we find that incarceration leads to short-term decreases in criminal justice contact. We find no evidence of longer-term impacts of exposure to incarceration. We refer the reader to Appendix G for details on our empirical approach and findings.

Our analysis has some limitations. First, incarceration may affect outcomes among subgroups that we are underpowered to detect (Aizer and Doyle, 2015; Mueller-Smith, 2015; Jordan, Karger, and Neal, 2023). Second, our empirical setting does not allow us to isolate the effects of long carceral sentences (e.g., five or ten years) vs noncarceral conviction. A higher "dosage" of incarceration may have more impact. Third, some people with noncarceral convictions could have been incarcerated prior to trial and thus may have already experienced some incarceration, reducing the difference in carceral exposure between those incarcerated post-trial and those receiving noncarceral convictions.

In a similar vein, some people who receive noncarceral conviction become incarcerated in the future, either because of new criminal convictions or because of probation violations. The difference in carceral exposure between these two groups thus becomes smaller over time. However, our evidence suggests that there remains a substantial difference in exposure to incarceration across these two groups. Appendix Figure E.2 shows how much "incarceration catch-up" occurs for those who receive noncarceral sentences compared to those who receive carceral sentences, both for new crimes and for technical violations resulting in probation revocation. These results suggest that, while there is some catch-up, more than 50% of those receiving a noncarceral sentence are not incarcerated over the next seven years.

Overall, the results from Sections *IV.B* and *IV.C* imply that incarceration's influence on the revolving door is limited, while noncarceral conviction may hold greater importance. Our findings on the effects of incarceration align with the recent literature review by Loeffler and Nagin (2022). Most of the studies that find incarceration to be criminogenic are looking at

pretrial detention, rather than post-sentencing incarceration. Since pretrial detention also increases the probability of conviction (Gupta, Hansman, and Frenchman, 2016; Leslie and Pope, 2017; Stevenson, 2018; Dobbie, Goldin, and Yang, 2018), these studies are effectively estimating the joint effect of conviction and incarceration. In contrast, most studies evaluating the impact of post-conviction incarceration do not find evidence of effects lasting beyond the incapacitation period. In line with our findings, most studies find no evidence that it is an important contributor to the revolving door.

IV.D Robustness and subgroup analyses

In this subsection, we provide a brief overview of robustness checks that we discuss in more detail in Appendix E.1. The results from the previous section are robust to the choice of sample restrictions and controls, as shown in Appendix Figures E.3–E.6. In particular, the results are similar when we drop specific crime types, for example drug cases, for which diversion is more likely to happen than for other offenses. Appendix Figures E.3–E.6 also show that the 2SLS estimates and standard errors remain similar when we more flexibly control for non-focal stringency.[30] Appendix Table E.9 shows that the results are robust to varying the definition of recidivism, and considering counts of new offenses and charges. Appendix E.4 shows that the results are robust to correcting for measurement error in stringency using Empirical Bayes methods. Additionally, Appendix Figure E.7 demonstrates no differential mobility out of Virginia based on incarceration outcomes.[31]

To examine treatment effect heterogeneity, we first break out the results based on whether a person has a prior felony conviction (Appendix Table E.6), since avoiding a first felony conviction might play an especially pivotal role. 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. Sample size limitations again preclude conclusive inference about heterogeneity in the impacts of incarceration across those with and without a recent felony, although point estimates are similar for the two groups.[32]

We additionally explore heterogeneity across race and zip code income level. We provide more details in Appendix E.1. We find qualitatively similar patterns across Black and non-Black defendants. We find suggestive evidence that the impacts of noncarceral conviction are larger for people living in zip codes with above-median poverty rates. Felony convictions might have greater consequences for poorer individuals, perhaps because convictions block access to housing, employment, or public assistance.

---

[30]See Table C.2, which provides further robustness to the choice of controls.

[31]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.

[32]We define our prior felony indicator as a prior felony within the last five years. Unlike Jordan, Karger, and Neal (2023), who can isolate first felony convictions using age restrictions, our data does not include age.

IV.E   Testing for and characterizing bias in the 2SLS results

In Section III.D, we showed that the 2SLS estimates may be asymptotically biased if the UPM assumption fails. In this subsection, we describe and implement an empirical test for this assumption. We then use theory and external evidence to discuss the likely magnitude and direction of the bias in our setting.

IV.E.1   Testing the UPM assumption

The UPM assumption has testable implications. If instrumental variation is only causing flows between two treatments, there should be no movement in or out of the third treatment. In our setting, this implies:

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

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

To build intuition for the first implication, consider those incarcerated in the ordered model from Section III.C. When holding incarceration stringency fixed, varying conviction stringency will move people between dismissal and conviction, but will not move people into or out of incarceration. If the instruments are treatment-specific, the observed characteristics of incarcerated individuals should not change. If the characteristics of incarcerated individuals do change, then there must be flows into and out of incarceration, which implies that the instrument is moving people across more than one margin. More generally, this would imply that $\text{UPM}(Z_c \mid Z_i)$ is violated, as the UPM assumption plus stringency instruments (and A1–A4) ensures compliers move across only one margin. A similar argument holds for the second testable implication.

These conditions allow us to adjudicate between models of judge decision-making introduced in Section III.C. In particular, (1) and (2) 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.[33] We test implication (1) by regressing predicted recidivism on the noncarceral 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. Table VI reports the results, where Panel A presents tests

---

[33]Predicted recidivism variables are created by regressing recidivism post-release if incarcerated, or post-conviction/dismissal otherwise, on offense type, socio-demographic controls, and month, court, and day-of-the-week fixed effects. Using these regressions, we construct measures of predicted recidivism within one year, two to four years, five to seven years, and within seven years after case disposition.

for (1) and Panel B tests for (2).[34] Appendix Table E.10 shows results for both tests using specific defendant characteristics (criminal record, offense type, and demographics) instead of predicted recidivism.

Using the predicted recidivism index, we reject $\text{UPM}(Z_c \mid Z_i)$ and $\text{UPM}(Z_i \mid Z_d)$, which 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. These results suggest the UPM assumption does not hold in our setting, and so our 2SLS estimates are potentially asymptotically biased.

IV.E.2   Sign and magnitude of asymptotic bias

Proposition 1 states that when UPM does not hold (but A1–A4 and CPM do), 2SLS estimands will be positively-weighted averages of the Wald estimands in equation (9). In this section, we demonstrate how the expression in equation (9) can be combined with theory and external evidence to reason about the direction and magnitude of bias in 2SLS estimands. We consider each margin of interest separately. Throughout this discussion, we will assume that CPM holds, as it does in each of the three judge decision-making models we considered. We also assume A1–A4 from Section III hold.

Noncarceral conviction vs dismissal. For simplicity, we discuss the bias term in the context of the special case where two judges have the same incarceration rate but differing rates of noncarceral conviction. In this case, equation (9) shows that the bias term in the Wald estimand is less than but proportional to $\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c}$, which is the difference in the impact of incarceration vs conviction between those near the incarceration-dismissal margin and those near the incarceration-conviction margin. This allows us to reason about the bias's sign and magnitude based on how incarceration may affect recidivism differently for these groups, compared to conviction. We separately consider the long- and short-run effects—with "long-run" referring to the post-incapacitation period.

Table VI shows that the average predicted recidivism rate among the incarcerated group increases in response to increasing Z_c while controlling for Z_i. This pattern suggests that individuals shifting from dismissal to incarceration exhibit higher predicted recidivism rates than those shifting from incarceration to conviction. This finding aligns with reasonable expectations about case characteristics close to the different decision margins. Cases near the incarceration-dismissal margin likely involve defendants with serious charges and/or extensive criminal histories ("high severity" cases), who would be incarcerated if sufficient evidence existed for conviction. In contrast, cases at the incarceration-conviction margin typically

---

[34]When implementing this test, we are maintaining other assumptions we make throughout the paper, such as the assumption that judge stringencies do not idiosyncratically depend on defendant characteristics and CPM. Results are similar when including flexible controls for the other stringency measure.

involve lower severity offenses but with stronger evidentiary support. Since criminal history is among the strongest predictors of recidivism, these high-severity cases at the incarceration-dismissal margin would likely present higher recidivism risk, consistent with our empirical bias test results. If, in the short run, incarceration affects recidivism primarily through incapacitation, then shifting prison beds towards people at a higher risk of recidivism will reduce recidivism ($\Delta_{d\to i}^{Y_i - Y_c} < \Delta_{i\to c}^{Y_i - Y_c}$). In this case, the bias term in equation (9) is negative.

However, the magnitude of the composition change shown in Table VI is relatively small: a ten percentage point increase in noncarceral conviction stringency increases one-year predicted recidivism among the incarcerated group by 0.1 percentage points. This suggests that either the proportion of $i \to c$ compliers is small, or the two groups have similar observable characteristics and therefore potentially similar treatment effects. Both would imply that the magnitude of the bias is small.

Turning to the long run (i.e., the post-incapacitation period), if incarceration *only* has incapacitation effects, then the impact of incarceration vs conviction is zero after the incapacitation period. In that case, the 2SLS estimates are asymptotically unbiased. Several pieces of evidence indicate that incarceration may not affect recidivism beyond the incapacitation period. In our setting, using an alternative research design, the regression-discontinuity evidence presented in Section IV.C demonstrates that incarceration reduces recidivism only in the short run (for a duration approximately matching the incapacitation period) among individuals on the margin of conviction and incarceration. This finding aligns with broader U.S. evidence, where most studies conclude that the impact of post-conviction incarceration on recidivism is negligible (Loeffler and Nagin, 2022; Garin et al., 2024).

While this evidence suggests limited effects beyond incapacitation, incarceration could theoretically influence recidivism through other channels. For instance, Jordan, Karger, and Neal (2023) finds that prison may serve as a stronger post-release deterrent for individuals with fewer prior convictions. These individuals, who likely have a lower propensity to recidivate, are disproportionately represented near the incarceration-conviction margin compared to the incarceration-dismissal margin.[35] In such cases, $\Delta_{d\to i}^{Y_i - Y_c} > \Delta_{i\to c}^{Y_i - Y_c}$, potentially introducing positive bias. However, two key findings in our setting suggest this upward bias is unlikely. First, our analysis reveals no evidence of heterogeneous treatment effects of incarceration by prior conviction status (Panel B of Appendix Table E.6), although these estimates lack precision. Second, as established earlier, the preponderance of empirical evidence (including evidence from our own empirical context) indicates that the post-incapacitation effects of incarceration vs conviction on recidivism are negligible.

Overall, the arguments above suggest that a violation of UPM would likely result in a modest negative bias in our 2SLS estimates of noncarceral conviction effects in the short run, with negligible bias in the long run. Hence, it is unlikely that our qualitative conclusions about the impact of noncarceral conviction vs dismissal would be overturned as a result of a

---

[35]Our implementation of test (1) using prior convictions instead of predicted recidivism reveals that $c \to i$ compliers have a lower prior conviction rate than $d \to i$ compliers (Appendix Table E.10).

violation of the UPM assumption.

Incarceration vs noncarceral conviction. As before, we discuss bias in the context of the simple case where two judges have the same noncarceral conviction rate but differing rates of incarceration. A derivation similar to the proof of Proposition 1 shows that the bias term for the impact of incarceration will be smaller than but proportional to $\Delta_{d \to i}^{Y_c - Y_d} - \Delta_{c \to d}^{Y_c - Y_d}$, and zero in the absence of $d \to i$ compliers. As evidenced by the results of our empirical test in Section *IV.E.1* , cases near the incarceration-dismissal margin have higher predicted recidivism (they are "high-severity" cases) than those on the conviction-dismissal margin.

Thus, to evaluate the bias for our estimates of the impact of incarceration vs conviction, we need to know whether a felony conviction (vs dismissal) will affect recidivism more for high-severity cases than for low-severity cases. One possibility is that a felony conviction increases recidivism more for low-severity cases, which seems reasonable because low-severity cases are less likely to have a prior felony on their criminal record and the first felony conviction is likely to have greater marginal impact than subsequent ones. If this conjecture is true, the bias term would be negatively signed. However, we find no discernible difference in the impact of conviction vs dismissal across crime types or priors (Appendix Tables E.4 - E.6). In addition, the compositional changes shown in Table VI and Appendix Table E.10 are relatively small. If the compositional shifts are minimal, then either the proportion of $d \to i$ compliers is small, or the $c \to d$ and $d \to i$ compliers have similar observable characteristics. In the first case, the bias would also be small. In the latter, the bias would be small if similar predicted recidivism implies similar treatment effects.

V   AN ALTERNATIVE APPROACH TO IDENTIFICATION AND ESTIMATION OF MARGIN-SPECIFIC TREATMENT EFFECTS

In the prior section, we found that our empirical test rejects the UPM assumption. Although we argued that the bias resulting from a violation of UPM is likely small given the specifics of our setting, this section explores alternative approaches based on assumptions that are not rejected by our test. This exercise serves as a robustness check on our qualitative conclusions and as a proof of concept for other researchers studying settings where examination of the expression in Section *III.D* raises more substantial concerns about bias.

Below we present a method for identifying margin-specific treatment effects in unordered choice settings. This approach does not require judge stringencies to satisfy UPM, nor does it place restrictions on treatment effect heterogeneity. Rather, we build on the approach in Mountjoy (2022), adopting the underlying assumptions. Furthermore, since this approach requires treatment-specific instruments, we suggest several alternative assumptions that can be used to construct such instruments.

V.A   Recovering treatment-specific instruments

Equation (7) sets up the unordered choice model and defines judge-specific thresholds (π_c and π_i), which are treatment-specific instruments if they can be exogenously varied while holding the other threshold fixed. The thresholds are not directly observed, but we observe the shares of cases ending in dismissal, conviction, and incarceration for each judge. We aim to recover π_c and π_i from the observed shares. Our setup has similarities to models in industrial organization where product shares are observed for different markets.[36] We therefore leverage identification results from the IO literature and adapt them to our context of judge decision-making. Berry, Gandhi, and Haile (2013) outline assumptions under which thresholds can be inverted from shares, and Berry and Haile (2024) show that judge-specific thresholds can be identified without invoking identification at infinity arguments.[37]

While these papers show that the thresholds are nonparametrically identified, we make additional parametric assumptions for tractability in estimation and show that the results are broadly similar under a set of parametric assumptions.[38] Our main specification assumes the unobserved shocks are each the sum of a shock with a standard logistic distribution and a random effect with a correlated multivariate normal distribution (η and ε in the equation below). We can then write the returns in the unordered choice model as

$$R_{ncj} = \beta_c - \pi_c^j + \gamma_c' X_n + \eta_{nc} + \varepsilon_{nc}$$
$$R_{nij} = \beta_i - \pi_i^j + \gamma_i' X_n + \eta_{ni} + \varepsilon_{ni},$$

where n represents the case, c and i indicate conviction or incarceration, j the judge, X_n are characteristics about the defendant or case, and R_{ncj} and R_{nij} represent the returns to a specific outcome for a specific case assigned to judge j. We assume $f(\varepsilon_{nc}, \varepsilon_{ni})$ has a standard logistic distribution and $g(\eta_{nc}, \eta_{ni}) \sim N(0, \Sigma)$.[39]

We estimate the model above by judicial circuit and three-year bin, which allows the model parameters to differ across circuits and over time. We then use the estimated model parameters to recover estimates of π_j^c and π_j^i.

---

[36] Unlike many applications in the industrial organization literature, our setting features quasi-random assignment of cases to judges, implying that the judge thresholds ($\pi_c(Z_c, Z_i)$ and $\pi_i(Z_c, Z_i)$) are independent of the characteristics of the case (V_c and V_i).

[37] Applying Berry and Haile (2024)'s identification argument to our setting requires the existence of three continuous covariates whose loadings do not vary across judges. See Appendix F.1 for details. Kamat, Norris, and Pecenco (2024) provide a partial identification strategy that does not require such conditions on covariates, but uses a sequential model combined with a "latent monotonicity" assumption, and recovers bounds rather than point estimates.

[38] Making additional parametric assumptions for estimation is common and often necessary in this literature. See Berry and Haile (2024) for a detailed discussion.

[39] We also consider two alternative specifications that are less flexible but easier to implement: (1) no random effect and the unobserved shocks follow standard logistic distributions and (2) Σ is a diagonal matrix.

V.B Recovering margin-specific effects

We refer to the newly constructed treatment-specific instruments—the estimated judge-specific thresholds—as $\tilde{Z}_c$ and $\tilde{Z}_i$, to distinguish them from the stringency instruments Z_c and Z_i.

Even with treatment-specific instruments, 2SLS estimands are difficult to interpret, as they are weighted averages of treatment effects that correspond to different margins, as visualized in Panel (c) of Figure III. In our context, shifting from $\tilde{z}_c$ to $\tilde{z}'_c$ while holding $\tilde{Z}_i$ fixed would yield a weighted average of the LATE for $i \to c$ compliers and the LATE for $d \to c$ compliers. The central objective of Mountjoy (2022) is to decompose the 2SLS estimand, obtained using a treatment-specific instrument, into two treatment-specific effects.

We closely follow Mountjoy (2022) in estimating the impacts on the two margins discussed above. This method relies on assumptions A1–A4, defined for $\tilde{z}_c$ and $\tilde{z}_i$, plus one additional assumption: "comparable compliers." This assumption requires that the $i \to c$ compliers from decreasing $\tilde{z}_i$ have the same potential outcome when convicted as $i \to c$ compliers from increasing $\tilde{z}_c$ at their limits (see Appendix F for a formal definition). Under this set of assumptions, Mountjoy (2022) shows how to identify and estimate $E[Y(c) - Y(d) \mid d \to c$ complier w.r.t $(\tilde{z}_c, \tilde{z}_i) \to (\tilde{z}'_c, \tilde{z}_i)]$ and $E[Y(i) - Y(c) \mid i \to c$ complier w.r.t $(\tilde{z}_c, \tilde{z}_i) \to (\tilde{z}'_c, \tilde{z}_i)]$. We follow Mountjoy (2022) in our approach and provide additional details in Appendix F.

While we do not invoke the UPM assumption in this section, we introduce additional assumptions in both the construction of treatment-specific instruments and in estimating treatment-specific effects.[40] The assumptions we consider in this section are not necessarily weaker or stronger than those supporting a causal interpretation of the 2SLS estimates.

V.C Results

Panel A of Table VII reports estimates for the noncarceral conviction vs dismissal margin. The point estimates are similar to the 2SLS estimates reported in Section IV, with somewhat smaller estimates for year 1 and somewhat larger estimates for later years. For example, the 2SLS estimate for a future felony charge within the first seven years is 0.24 (95% CI: 0.05–0.43), while the estimate from this alternative approach is 0.27 (95% CI: 0.10,0.47). For ten of the 12 estimates, the 2SLS estimates fall within the confidence interval of the new estimates. Panel B reports estimates for the incarceration vs noncarceral conviction margin. Again, these results are similar to the 2SLS estimates, with somewhat smaller estimates in year 1.[41]

Overall, the 2SLS estimates and the estimates based on this alternative approach tell a similar story: noncarceral conviction increases future criminal justice contact in the long run, while for incarceration the evidence only supports short-term incapacitation effects.

---

[40]For identification, we assume the unordered model, "comparable compliers," and the existence of additive covariates whose loadings do not vary across judges. For estimation, we additionally make distributional assumptions about the error terms.

[41]We include additional results under the two alternative assumptions outlined in footnote 39 in Appendix Tables F.1 and F.2.

VI   CONCLUSION

We study the role of noncarceral conviction in driving future criminal justice contact, and compare it to the role of incarceration. Our analyses consistently demonstrate that noncarceral conviction increases future criminal justice contact (relative to dismissal). In contrast, our analysis of incarceration (relative to noncarceral conviction) only finds evidence for a shorter-term decrease in recidivism, which coincides with the typical period of incapacitation. While our findings support the existence of a "revolving door" effect in the criminal justice system, it primarily operates through the channel of noncarceral conviction rather than through incarceration.

We also discuss methodological challenges stemming from multiple treatment alternatives in the commonly-used random judge research design. We develop an empirical framework for interpreting 2SLS estimands using judge stringency instruments under heterogeneous treatment effects. We provide assumptions that allow the estimands to be interpreted as causal and margin-specific. In particular, we show that requiring judge instruments to be treatment-specific is sufficient (in addition to straightforward extensions of exclusion, random assignment, relevance, and rich controls). We characterize models of judge decision-making that are consistent with treatment specificity, and propose an empirical test for this assumption. For cases where treatment specificity fails, we derive an expression for the asymptotic bias, enabling researchers to assess likely direction and magnitude of bias using features of the institutional setting, theoretical arguments, or prior empirical evidence. Finally, we propose and implement an empirical approach that allows us to identify causal and margin-specific treatment effects under an alternative set of assumptions that does not include treatment-specificity.

Several policy approaches could reduce either the frequency of felony convictions or their lasting impact. The number of felony convictions could be reduced by expanding felony diversion programs, decriminalizing certain offenses, or downgrading certain charges to misdemeanors. To diminish the impact of existing felony convictions, policymakers could limit the accessibility or permissible uses of criminal records. For instance, limiting how long criminal records are publicly available could mitigate the employment effects of having a criminal record, potentially reducing recidivism by increasing formal employment options (Cullen, Dobbie, and Hoffman, 2023). Likewise, reducing automatic escalations within the penal system, such as charge upgrades or sentence enhancements for those with a prior felony conviction, could mitigate the impact of a criminal record (Rose, 2021).

Our analysis suggests that such reforms could help address the penal system's revolving door problem. While other policy considerations remain relevant—including legitimate uses of felony conviction records in hiring decisions and sentencing—the scale of felony convictions in the U.S. demands careful attention to their downstream effects. With an estimated 8% of adults and 33% of Black adult men holding felony conviction records (Shannon et al., 2017), the impact of these convictions on future criminal justice contact should be an important part of policy discussions.

REFERENCES

Agan, Amanda Y., Jennifer L. Doleac, and Anna Harvey (2023). "Misdemeanor prosecution". *The Quarterly Journal of Economics* 138.3, pp. 1453–1505.

Agan, Amanda Y., Matthew Freedman, and Emily Owens (2021). "Is your lawyer a lemon? Incentives and selection in the public provision of criminal defense". *Review of Economics and Statistics* 103.2, pp. 294–309.

Agan, Amanda Y., Andrew Garin, Dmitri Koustas, Alexander Mas, and Crystal Yang (2024a). "Can you Erase the Mark of a Criminal Record? Labor Market Impacts of Criminal Record Remediation". Tech. rep. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w32394/w32394.pdf.

— (Sept. 2024b). "The Labor Market Impacts of Reducing Felony Convictions". *American Economic Review: Insights* 6.3, pp. 341–58.

Agan, Amanda Y. and Sonja Starr (2018). "Ban the box, criminal records, and racial discrimination: A field experiment". *The Quarterly Journal of Economics* 133.1, pp. 191–235.

Aizer, Anna and Joseph J. Jr. Doyle (May 2015). "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges". *Quarterly Journal of Economics* 130.2. MAG ID: 2118051501, pp. 759–803.

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

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

Arnold, David, Will Dobbie, and Peter Hull (Sept. 2022). "Measuring Racial Discrimination in Bail Decisions". *American Economic Review* 112.9, pp. 2992–3038.

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

Augustine, Elsa, Johanna Lacoe, Steven Raphael, and Alissa Skog (2022). "The impact of felony diversion in San Francisco". *Journal of Policy Analysis and Management* 41.3, pp. 683–709.

Avi-Itzhak, Benjamin and Reuel Shinnar (1973). "Quantitative models in crime control". *Journal of Criminal Justice* 1.3, pp. 185–217.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen (Feb. 2009). "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections*". *The Quarterly Journal of Economics* 124.1, pp. 105–147.

Berry, Steven T., Amit Gandhi, and Philip A. Haile (2013). "Connected Substitutes and Invertibility of Demand". *Econometrica* 81.5, pp. 2087–2111.

Berry, Steven T. and Philip A. Haile (2024). "Nonparametric Identification of Differentiated Products Demand Using Micro Data". *Econometrica* 92.4, pp. 1135–1162.

Bhuller, Manudeep, Gordon Dahl, Katrine Løken, and Magne Mogstad (2020). "Incarceration, recidivism, and employment". *Journal of Political Economy* 128.4, pp. 1269–1324.

Bhuller, Manudeep and Henrik Sigstad (2024). "2SLS with multiple treatments". *Journal of Econometrics* 242.1, p. 105785.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky (2022). "When is TSLS actually late?" Tech. rep. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w29709/w29709.pdf.

Blevins, Kristie R., Shelley Johnson Listwan, Francis T. Cullen, and Cheryl Lero Jonson (2010). "A general strain theory of prison violence and misconduct: An integrated model of inmate behavior". *Journal of Contemporary Criminal Justice* 26.2, pp. 148–166.

Chaisemartin, Clement de (2017). "Tolerating Defiance? Local Average Treatment Effects without Monotonicity". *Quantitative Economics* 8.2, pp. 367–96.

Chan, David C., Matthew Gentzkow, and Chuan Yu (2022). "Selection with Variation in Diagnostic Skill: Evidence from Radiologists". *Quarterly Journal of Economics* 137.2, pp. 729–83.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (Sept. 2014). "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood". *American Economic Review* 104.9, pp. 2633–79.

Chiricos, T., K. Barrick, W. Bales, and S. Bontrager (2007). "The labeling of convicted felons and its consequences for recidivism." *Criminology: An Interdisciplinary Journal* 45.3, pp. 547–581.

Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk (2024). "Eviction and poverty in American cities: Evidence from Chicago and New York". *Quarterly Journal of Economics*.

Craigie, Terry-Ann (2020). "Ban the box, convictions, and public employment". *Economic Inquiry* 58.1, pp. 425–445.

Cullen, Zoë, Will Dobbie, and Mitchell Hoffman (Feb. 2023). "Increasing the Demand for Workers with a Criminal Record". *The Quarterly Journal of Economics* 138.1, pp. 103–150.

Dahl, Gordon, Andreas Ravndal Kostøl, and Magne Mogstad (Aug. 2014). "Family Welfare Cultures". *The Quarterly Journal of Economics* 129.4, pp. 1711–1752.

Deshpande, Manasi and Michael Mueller-Smith (2022). "Does welfare prevent crime? The criminal justice outcomes of youth removed from SSI". *The Quarterly Journal of Economics* 137.4, pp. 2263–2307.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang (2018). "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges". *American Economic Review* 108.2, pp. 201–240.

Dobbie, Will and Jae Song (2015). "Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection". *American economic review* 105.3, pp. 1272–1311.

Doleac, Jennifer L. (2023). "Encouraging Desistance from Crime". *Journal of Economic Literature* 61.2, pp. 383–427.

Doyle, Joseph J. Jr. (2008). "Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care". *Journal of political Economy* 116.4, pp. 746–770.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova (2009). "The deterrent effects of prison: Evidence from a natural experiment". *Journal of political Economy* 117.2, pp. 257–280.

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai (2019). "Using a probabilistic model to assist merging of large-scale administrative records". *American Political Science Review* 113.2, pp. 353–371.

Farrar-Owens, Meredith (2013). "The evolution of sentencing guidelines in Virginia: An example of the importance of standardized and automated felony sentencing data". *Federal Sentencing Reporter* 25.3, pp. 168–170.

Finlay, Keith, Matthew Gross, Carl Lieberman, Elizabeth Luh, and Michael Mueller-Smith (Dec. 2024). "The Impact of Criminal Financial Sanctions: A Multistate Analysis of Survey and Administrative Data". *American Economic Review: Insights* 6.4, pp. 490–508.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie (Jan. 2023). "Judging Judge Fixed Effects". en. *American Economic Review* 113.1, pp. 253–277.

Garin, Andrew, Dmitri K Koustas, Carl McPherson, Samuel Norris, Matthew Pecenco, Evan K Rose, Yotam Shem-Tov, and Jeffrey Weaver (July 2024). "The Impact of Incarceration on Employment, Earnings, and Tax Filing". Working Paper 32747. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w32747/w32747.pdf.

Goldsmith-Pinkham, Paul, Maxim Pinkovskiy, and Jacob Wallace (2023). "The great equalizer: Medicare and the geography of consumer financial strain". Tech. rep. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w31223/w31223.pdf.

Graef, Lindsay, Sandra Mayson, Aurelie Ouss, and Megan T. Stevenson (2023). "Systemic Failure to Appear in Court". *U. Pa. L. Rev.* 172, p. 1.

Gross, Max and E Jason Baron (2022). "Temporary stays and persistent gains: The causal effects of foster care". *American Economic Journal: Applied Economics* 14.2, pp. 170–99.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman (2016). "The Heavy Costs of High Bail: Evidence from Judge Randomization". *The Journal of Legal Studies* 45.2, pp. 471–505.

Hagan, John (1993). "The social embeddedness of crime and unemployment". *Criminology* 31.4, pp. 465–491.

Heckman, James J., John Eric Humphries, and Gregory Veramendi (2016). "Dynamic treatment effects". *Journal of Econometrics* 191.2. Innovations in Measurement in Economics and Econometrics, pp. 276–292.

Heckman, James J. and Rodrigo Pinto (2018). "Unordered Monotonicity". *Econometrica* 86.1, pp. 1–35.

Heckman, James J., Sergio Urzua, and Edward Vytlacil (Aug. 2006). "Understanding Instrumental Variables in Models with Essential Heterogeneity". *The Review of Economics and Statistics* 88.3, pp. 389–432.

Heinesen, Eskil, Christian Hvid, Lars Johannessen Kirkebøen, Edwin Leuven, and Magne Mogstad (Oct. 2022). "Instrumental Variables with Unordered Treatments: Theory and Evidence from Returns to Fields of Study". NBER Working Paper 30574. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w30574/w30574.pdf.

Holzer, Harry J, Steven Raphael, and Michael A Stoll (2007). "The effect of an applicant's criminal history on employer hiring decisions and screening practices: Evidence from Los Angeles". *Barriers to reentry* 4.15, pp. 117–150.

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". *Econometrica* 62.2, pp. 467–475.

Imbens, Guido W. and Stefan Wager (2019). "Optimized regression discontinuity designs". *Review of Economics and Statistics* 101.2, pp. 264–278.

Irankunda, Armel, Gregory N. Price, Norense E. Uzamere, and Miesha J. Williams (2020). "Ex-Incarceree/Convict Status: Beneficial for Self-Employment and Entrepreneurship?" *The American Economist* 65.1, pp. 144–162.

IRS data (2005). SOI Tax Stats - Individual income tax statistics - ZIP Code data (SOI).

Jordan, Andrew, Ezra Karger, and Derek Neal (Dec. 2023). "Heterogeneous Impacts of Sentencing Decisions". Working Paper 31939. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w31939/w31939.pdf.

Kamat, Vishal, Samuel Norris, and Matthew Pecenco (2024). "Conviction, Incarceration, and Policy Effects in the Criminal Justice System". *Available at SSRN*. eprint: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4777635.

Kane, Thomas J and Douglas O Staiger (Dec. 2008). "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation". Working Paper 14607. National Bureau of Economic Research.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad (2016). "Field of Study, Earnings, and Self-Selection". *The Quarterly Journal of Economics* 131.3, pp. 1057–1112.

Kline, Patrick and Christopher R. Walters (July 2016). "Evaluating Public Programs with Close Substitutes: The Case of Head Start". *The Quarterly Journal of Economics* 131.4, pp. 1795–1848.

Kohler-Hausmann, Issa (2018). *Misdemeanorland: Criminal courts and social control in an age of broken windows policing.* Princeton University Press.

Kolesár, Michal and Christoph Rothe (2018). "Inference in regression discontinuity designs with a discrete running variable". *American Economic Review* 108.8, pp. 2277–2304.

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

Lee, Sokbae and Bernard Salanié (2018). "Identifying Effects of Multivalued Treatments". en. *Econometrica* 86.6, pp. 1939–1963.

Leslie, Emily and Nolan G. Pope (2017). "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments". *The Journal of Law and Economics* 60.3, pp. 529–557.

Lieberman, Carl, Elizabeth Luh, and Michael Mueller-Smith (2023). *Criminal court fees, earnings, and expenditures: A multi-state RD analysis of survey and administrative data.* US Census Bureau, Center for Economic Studies.

Loeffler, Charles E. and Daniel S. Nagin (2022). "The impact of incarceration on recidivism". *Annual Review of Criminology* 5, pp. 133–152.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand (2013). "Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt". *American economic review* 103.5, pp. 1797–1829.

Mayson, Sandra and Megan T. Stevenson (Mar. 2020). "Misdemeanors by the Numbers". *Boston College Law Review* 61.3.

Morris, Carl N. (1983). "Parametric Empirical Bayes Inference: Theory and Applications". *Journal of the American Statistical Association* 78.381, pp. 47–55.

Mountjoy, Jack (2022). "Community colleges and upward mobility". *American Economic Review* 112.8, pp. 2580–2630.

Mueller-Smith, Michael (2015). "The Criminal and Labor Market Impacts of Incarceration: Identifying Mechanisms and Estimating Household Spillovers". Working Paper. University of Michigan. eprint: https://sites.lsa.umich.edu/mgms/wp-content/uploads/sites/283/2015/09/incar.pdf.

Mueller-Smith, Michael, Benjamin Pyle, and Caroline Walker (Aug. 2023). "Estimating the Impact of the Age of Criminal Majority: Decomposing Multiple Treatments in a Regression Discontinuity Framework". Working Paper 31523. National Bureau of Economic Research. eprint: https://www.nber.org/system/files/working_papers/w31523/w31523.pdf.

Mueller-Smith, Michael and Kevin T. Schnepel (2021). "Diversion in the criminal justice system". *The Review of Economic Studies* 88.2, pp. 883–936.

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

Norris, Samuel (2019). "Examiner inconsistency: Evidence from refugee appeals". *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2018-75. eprint: https://bfi.uchicago.edu/wp-content/uploads/BFI_WP_201875.pdf.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver (Sept. 2021). "The Effects of Parental and Sibling Incarceration: Evidence from Ohio". *American Economic Review* 111.9, pp. 2926–63.

Pager, Devah (2003). "The mark of a criminal record". *American journal of sociology* 108.5, pp. 937–975.

Pager, Devah, Rebecca Goldstein, Helen Ho, and Bruce Western (2022). "Criminalizing Poverty: The Consequences of Court Fees in a Randomized Experiment". *American Sociological Review* 87.3, pp. 529–553.

PASC (2013). "County Time Served and Revocations: 2013 report". Tech. rep. Pennsylvania Commission on Sentencing.

Phelps, Michelle S. (2013). "The paradox of probation: Community supervision in the age of mass incarceration". *Law & policy* 35.1-2, pp. 51–80.

— (2017). "Mass probation: Toward a more robust theory of state variation in punishment". *Punishment & society* 19.1, pp. 53–73.

Philippe, Arnaud (2024). "Learning by doing. How do criminals learn about criminal law?" *American Economic Journal: Economic Policy* 16.3, pp. 27–60.

Rivera, R (2023). "Release, detain or surveil? The effects of electronic monitoring on defendant outcomes". *Unpublished manuscript, Columbia University*.

Rose, Evan (2021). "Who gets a second chance? Effectiveness and equity in supervision of criminal offenders". *The Quarterly Journal of Economics* 136.2, pp. 1199–1253.

Sampat, Bhaven and Heidi L. Williams (2019). "How do patents affect follow-on innovation? Evidence from the human genome". *American Economic Review* 109.1, pp. 203–36.

Shannon, Sarah K. S., Christopher Uggen, Jason Schnittker, Melissa Thompson, Sara Wakefield, and Michael Massoglia (2017). "The Growth, Scope, and Spatial Distribution of People With Felony Records in the United States, 1948-2010". *Demography* 54.5, pp. 1795–1818.

Sigstad, Henrik (2023). "Monotonicity among Judges: Evidence from Judicial Panels and Consequences for Judge IV Designs". *Working Paper. Available at SSRN.* eprint: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4534809.

Stevenson, Megan T. (2017). "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails". *The Review of Economics and Statistics* 99.5, pp. 824–838.

— (2018). "Assessing risk assessment in action". *Minn. L. Rev.* 103, p. 303.

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

Virginia Criminal Sentencing Commission (2021). Dataset "Sentencing Data FY2021".

Virginia's Circuit Courts (2021). Dataset from Circuit Criminal Court Cases (2000, 2020). Ben Schoenfeld conducted web scraping of court records and subsequently made the corresponding data publicly available for download. We submitted a request for the un-anonymized datasets, which are exclusively accessible to journalists, non-profit organizations, research institutions, or government bodies. Alternatively, anonymized data can be directly downloaded for each year spanning from 2000 to 2023 under the dataset category labeled "Circuit Criminal Court Cases".

Virginia's District Courts (2021). Dataset from District Criminal Court Cases Court Cases (2000, 2020). Ben Schoenfeld conducted web scraping of court records and subsequently made the corresponding data publicly available for download. We submitted a request for the un-anonymized datasets, which are exclusively accessible to journalists, non-profit organizations, research institutions, or government bodies. Alternatively, anonymized data can be directly downloaded for each year spanning from 2000 to 2023 under the dataset category labeled "District Criminal Court Cases".

Wolff, Nancy, Cynthia L. Blitz, Jing Shi, Jane Siegel, and Ronet Bachman (2007). "Physical violence inside prisons: Rates of victimization". *Criminal justice and behavior* 34.5, pp. 588–599.

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

Table I: Summary statistics

	Dismissed	Convicted	Incarcerated
	(1)	(2)	(3)
Offenses			
Drugs	0.35	0.33	0.29
Larceny	0.17	0.29	0.25
Assault	0.19	0.08	0.18
Fraud	0.09	0.16	0.10
Traffic	0.04	0.05	0.13
Burglary	0.06	0.07	0.08
Robbery	0.05	0.02	0.06
Sexual assault	0.03	0.02	0.03
Kidnapping	0.03	0.01	0.02
Murder	0.01	0.00	0.01
Defendant characteristics			
Black	0.57	0.51	0.60
Female	0.22	0.32	0.16
% of ppl in zip earning <25K	0.46	0.44	0.46
Has misdemeanor	0.06	0.09	0.08
Prior conviction within 5 years	0.14	0.10	0.22
Sentencing			
Incarceration length	0.00	0.00	27.45
Probation length	0.00	31.50	39.34
Median incar. leng.	0	0	12
Median prob. leng.	0	12	12
Percent of sample	16	30	55
Observations	28,589	54,640	100,152

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 noncarceral conviction. The final column shows estimates for those whose cases ended with incarceration. The incarceration and probation length medians and means are in months. Probation length is top-coded at 20 years. Our primary data source is Virginia Circuit Court Records from 2000-2012.

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

	Non-carceral conviction			Incarceration		
	(1)	(2)	(3)	(4)	(5)	(6)
Conviction stringency	0.63***	0.60***	0.59***			
	(0.033)	(0.032)	(0.046)			
Incarceration stringency			-0.011	0.62***	0.59***	0.60***
			(0.041)	(0.033)	(0.032)	(0.035)
Dismissal stringency						0.032
						(0.051)
Controls	No	Yes	Yes	No	Yes	Yes
Mean dep. var.	0.298	0.298	0.298	0.546	0.546	0.546
F-stat	360.3	339.5	165.3	346.7	350.7	287.8
N	183,381	183,381	183,381	183,381	183,381	183,381

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 noncarceral conviction. The first column includes the instrument along with court-by-year fixed effects, court-by-month fixed effects, and day-of-week fixed effects. The second column adds individual and case-level controls, and the third column adds the leave-one-out judge incarceration stringency. Columns (4)-(6) repeat this analysis, but with incarceration as the outcome; the final column controls for judge dismissal stringency. Standard errors are clustered at the judge-year level. Stars denote * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table III: Balance

	Convicted	Conv. stringency	Incarceration	Incar. stringency
	(1)	(2)	(3)	(4)
Any prior conv.	-0.1380***	-0.0000	0.1704***	0.0002
	(0.0029)	(0.0002)	(0.0032)	(0.0002)
Female	0.1207***	-0.0003*	-0.1241***	0.0002
	(0.0032)	(0.0002)	(0.0031)	(0.0002)
Black	-0.0413***	0.0002	0.0457***	-0.0002
	(0.0025)	(0.0002)	(0.0026)	(0.0002)
Has misdemeanor	0.0434***	0.0001	-0.0148***	0.0003
	(0.0047)	(0.0003)	(0.0050)	(0.0003)
Drugs	-0.0282***	0.0003	0.0705***	-0.0000
	(0.0037)	(0.0002)	(0.0041)	(0.0003)
Larceny	-0.0093***	0.0003	0.0995***	0.0003
	(0.0035)	(0.0002)	(0.0037)	(0.0002)
Assault	-0.1542***	-0.0011***	0.1576***	0.0012***
	(0.0035)	(0.0002)	(0.0043)	(0.0003)
Fraud	0.0251***	0.0004	0.0515***	0.0006*
	(0.0040)	(0.0003)	(0.0042)	(0.0003)
Traffic	-0.1858***	-0.0003	0.3307***	0.0006*
	(0.0042)	(0.0003)	(0.0048)	(0.0004)
Burglary	-0.0408***	-0.0001	0.0782***	0.0005
	(0.0043)	(0.0003)	(0.0047)	(0.0003)
Robbery	-0.0949***	-0.0002	0.1647***	0.0004
	(0.0048)	(0.0004)	(0.0059)	(0.0004)
Sexual assault	-0.1681***	-0.0007	0.2070***	0.0012**
	(0.0062)	(0.0005)	(0.0074)	(0.0006)
Kidnapping	-0.0631***	-0.0005	-0.0023	0.0006
	(0.0066)	(0.0006)	(0.0085)	(0.0006)
Murder	-0.1537***	-0.0012	0.1762***	0.0010
	(0.0076)	(0.0008)	(0.0119)	(0.0010)
F-stat joint F-test	574.563	3.759	820.355	2.652
P-value joint F-test	0.000	0.000	0.000	0.001
Observations	183,381	183,381	183,381	183,381

Note: This table shows estimates from regressions of either case outcomes (noncarceral conviction or incarceration indicators) or judge stringency measures on case characteristics. Regressions include court-by-year fixed effects, court-by-month 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. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01. To see the balance table in standard deviation units, see Appendix Table D.1

Table IV: Noncarceral conviction and recidivism

	Year 1		Years 2-4		Year 5-7		Year 1-7	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Fut. charge	-0.002	0.105**	0.004	0.087	0.006**	0.078	0.011***	0.235**
	(0.002)	(0.046)	(0.003)	(0.075)	(0.002)	(0.075)	(0.004)	(0.097)
Fut. conviction	0.001	0.135***	0.008***	0.115	0.007***	0.055	0.014***	0.300***
	(0.002)	(0.043)	(0.003)	(0.073)	(0.002)	(0.071)	(0.004)	(0.095)
Fut. incarceration	0.001	0.113***	0.006**	0.059	0.005**	-0.024	0.012***	0.215***
	(0.002)	(0.037)	(0.002)	(0.063)	(0.002)	(0.057)	(0.003)	(0.083)
Ctrl. comp. mean: fut. chrg.	0.157	0.157	0.301	0.301	0.237	0.237	0.493	0.493
Ctrl. mean: fut. chrg.	0.089	0.089	0.170	0.170	0.129	0.129	0.297	0.297
Ctrl. comp. mean: fut. conv.	0.137	0.137	0.263	0.263	0.226	0.226	0.459	0.459
Ctrl. mean: fut. conv.	0.076	0.076	0.148	0.148	0.114	0.114	0.268	0.268
Ctrl. comp. mean: fut. incar.	0.135	0.135	0.287	0.287	0.276	0.276	0.522	0.522
Ctrl. mean: fut. incar.	0.054	0.054	0.109	0.109	0.083	0.083	0.204	0.204
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table shows regression estimates of the impact of noncarceral conviction on recidivism, measured by any future felony charge, any future felony conviction, and any future felony incarceration. Recidivism is defined relative to the date of sentencing within the time window shown at the top of each column: 1 year, 2-4 years, 5-7 years and up to 7 years. For each outcome, we report ordinary least squares (OLS) and instrumental variable (IV) estimates. All IV regressions control for judge incarceration stringency. For the OLS estimates, we regress recidivism on having a conviction (regardless of incarceration status), controlling for incarceration. The estimates presented are the coefficient on the conviction variable. The middle portion of the table reports the control complier mean and control mean for each of the outcomes we consider. Control means are calculated for cases that end in dismissal. See Appendix E.5 for details on the calculation of control complier means. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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 V: Incarceration and recidivism

	Year 1		Years 2-4		Years 5-7		Years 1-7	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Fut. charge	-0.022***	-0.097***	0.013***	-0.016	0.025***	0.004	0.022***	-0.071
	(0.002)	(0.029)	(0.002)	(0.047)	(0.002)	(0.040)	(0.003)	(0.059)
Fut. conviction	-0.018***	-0.112***	0.013***	-0.038	0.023***	0.021	0.022***	-0.106*
	(0.001)	(0.029)	(0.002)	(0.047)	(0.002)	(0.039)	(0.003)	(0.058)
Fut. incarceration	-0.010***	-0.071***	0.017***	0.007	0.021***	0.052	0.027***	-0.029
	(0.001)	(0.024)	(0.002)	(0.041)	(0.002)	(0.032)	(0.003)	(0.051)
Ctrl. comp. mean: fut. chrg.	0.122	0.122	0.199	0.199	0.147	0.147	0.370	0.370
Ctrl. mean: fut. chrg.	0.088	0.088	0.175	0.175	0.132	0.132	0.306	0.306
Ctrl. comp. mean: fut. conv.	0.084	0.084	0.168	0.168	0.113	0.113	0.311	0.311
Ctrl. mean: fut. conv.	0.077	0.077	0.159	0.159	0.120	0.120	0.283	0.283
Ctrl. comp. mean: fut. incar.	0.043	0.043	0.071	0.071	0.051	0.051	0.166	0.166
Ctrl. mean: fut. incar.	0.055	0.055	0.115	0.115	0.084	0.084	0.212	0.212
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table shows regression estimates of the impact of incarceration on recidivism, measured by any future felony charge, any future felony conviction, and any future felony incarceration. Recidivism is defined relative to the date of sentencing within the time window shown at the top of each column: 1 year, 2-4 years, 5-7 years and up to 7 years. For each outcome, we report ordinary least squares (OLS) and instrumental variable (IV) estimates. 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). The middle portion of the table reports the control complier mean and control mean for each of the outcomes we consider. Control means are calculated for cases that end in noncarceral conviction. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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 VI: Testing for treatment-specificity

	Predicted recidivism			
	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: UPM$(Z_c \mid Z_i)$ – ordered model				
Conviction stringency (Z_c)	0.013***	0.030***	0.023***	0.048***
	(0.0039)	(0.0092)	(0.0072)	(0.014)
Mean dep. var.	0.093	0.202	0.153	0.346
N	100,152	100,152	100,152	100,152
Panel B: UPM$(Z_i \mid Z_d)$ – sequential and ordered model				
Incarceration stringency (Z_i)	-0.012***	-0.026**	-0.019**	-0.040**
	(0.0044)	(0.010)	(0.0082)	(0.017)
Mean dep. var.	0.090	0.183	0.138	0.320
N	28,589	28,589	28,589	28,589

Note: This table tests for treatment specificity following the method outlined in Section *IV.E.1* . 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. 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, 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 VII: Margin-specific treatment effects: an alternative approach

	Mixed logit with correlated normal random effects			
	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Noncarceral conviction vs dismissal (C vs D)				
Felony charge:	0.080**	0.209***	0.145***	0.269***
	[0.002,0.164]	[0.093,0.375]	[0.030,0.302]	[0.097,0.472]
	{0.073}	{0.148}	{0.127}	{0.305}
Felony conviction:	0.086**	0.225***	0.115**	0.317***
	[0.010,0.169]	[0.108,0.355]	[0.018,0.229]	[0.159,0.497]
	{0.060}	{0.124}	{0.126}	{0.252}
Felony incarceration:	0.062*	0.161***	0.075*	0.223***
	[-0.002,0.129]	[0.041,0.286]	[-0.012,0.169]	[0.065,0.412]
	{0.055}	{0.104}	{0.092}	{0.237}
Panel B: Incarceration vs noncarceral conviction (I vs C)				
Felony charge:	-0.047***	0.039	0.001	-0.033
	[-0.076,-0.011]	[-0.018,0.101]	[-0.058,0.059]	[-0.117,0.058]
	{0.081}	{0.163}	{0.139}	{0.325}
Felony conviction:	-0.039**	0.033	0.009	-0.028
	[-0.068,-0.008]	[-0.024,0.086]	[-0.047,0.062]	[-0.110,0.059]
	{0.073}	{0.150}	{0.124}	{0.300}
Felony incarceration:	-0.015	0.048*	0.011	-0.031
	[-0.041,0.011]	[-0.005,0.095]	[-0.037,0.058]	[-0.111,0.059]
	{0.052}	{0.097}	{0.092}	{0.229}
Controls	Yes	Yes	Yes	Yes

This table reports margin-specific estimates of the impact of noncarceral conviction vs dismissal (Panel A) and incarceration vs noncarceral conviction (Panel B) using an unordered multinomial model based on the methodology developed in Mountjoy (2022). The treatment-specific instruments are recovered as described in Section V.A using a mixed-logit specification for the choice model where the intercept includes a correlated multivariate normal random effect and controls for female and Black indicators, an indicator for whether any charges are for violent crimes, an indicator for whether any charges are for property crimes, and indicator for whether any charges are for drug crimes, the number of charges, the time since last offense, and the number of misdemeanor charges associated with the case. The choice model is fit by circuit and three-year time window. The estimates then use the recovered treatment-specific instruments in the method developed by Mountjoy (2022), where we include the same controls plus circuit and year fixed effects. The curly brackets report control-group complier means. In the top panel, this is the mean outcome for compliers whose cases were dismissed, while for the bottom panel, it is for those convicted but not incarcerated. 95% confidence intervals are reported in brackets and are based on 500 bootstrap samples. Stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on the 90%, 95%, and 99% confidence intervals.

Figure I: Ordered choice model

(a)
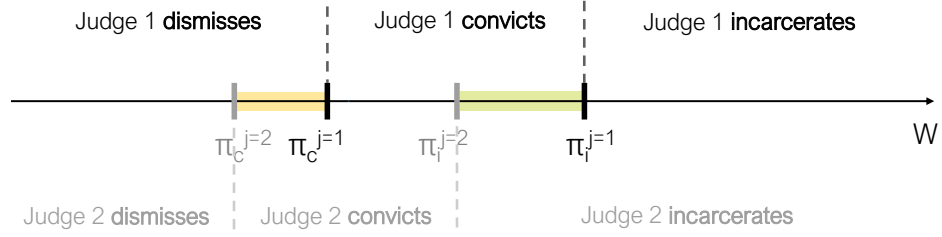
(b)
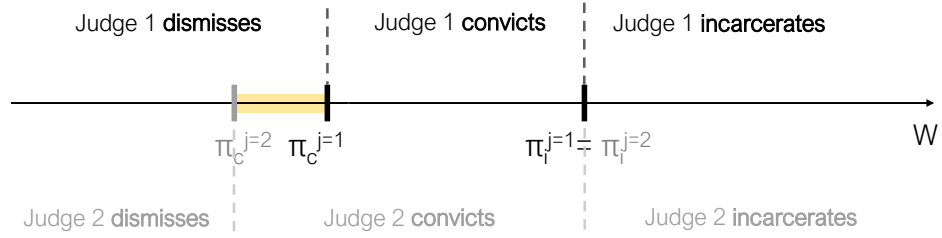
Note: This figure visualizes how, under the ordered choice model discussed in Section *III.C.1* , judges classify individuals into incarceration, conviction, and dismissal depending on the cases' unobservable W. Panel (a) visualizes this for two arbitrary judges, and Panel (b) does so for two judges with the same incarceration stringency but different conviction stringencies.

Figure II: Sequential choice model

(a)

(b)

(c)
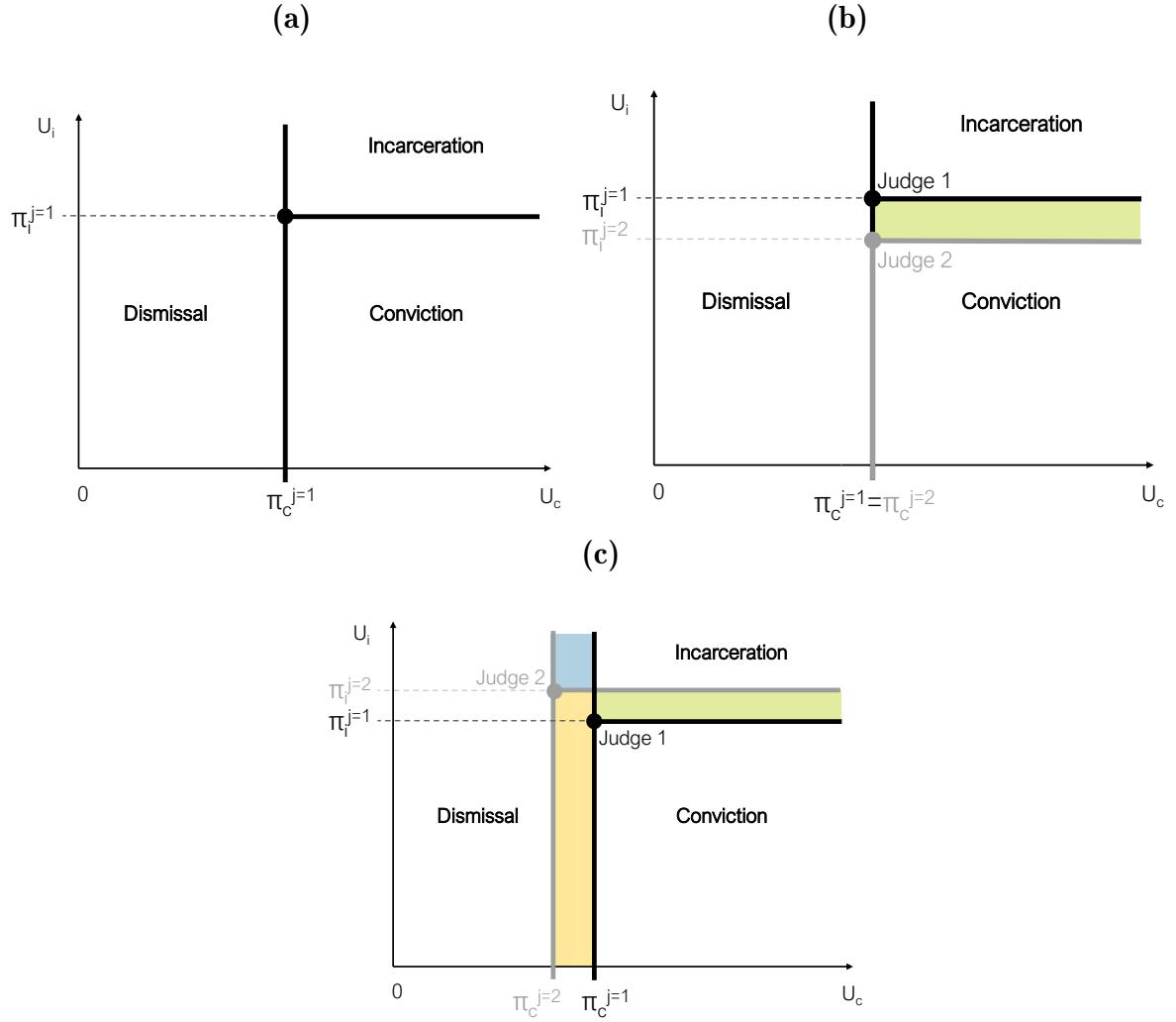
Note: This figure visualizes how, under the sequential choice model discussed in Section *III.C.2* , judges classify individuals into incarceration, conviction, and dismissal depending on the cases' unobservable U_i and U_c. Panel (a) visualizes this for an arbitrary judge, Panel (b) does so for two 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 a higher conviction stringency.

Figure III: Unordered multinomial choice model

(a)

(b)

(c)
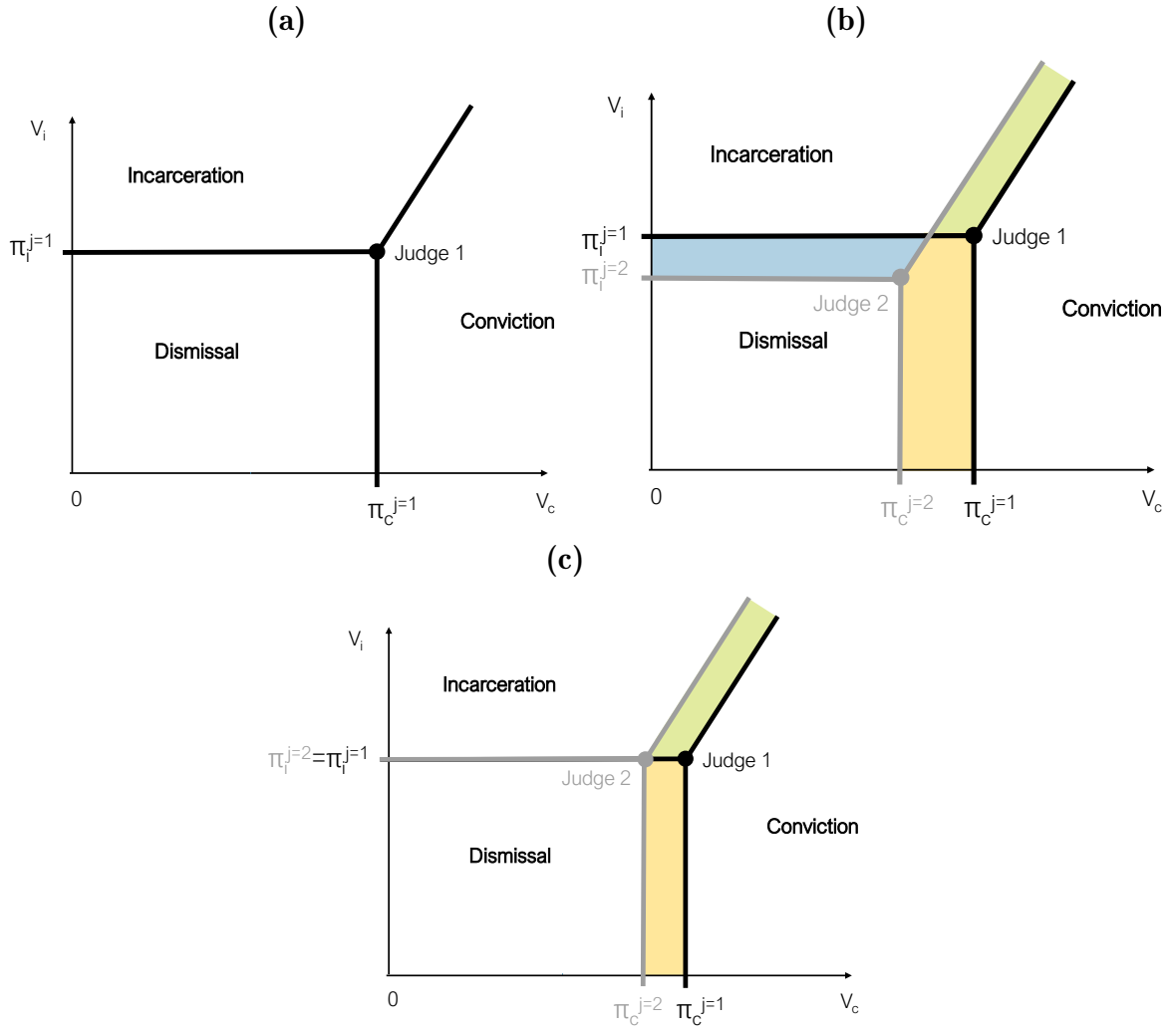
Note: This figure visualizes how, under the unordered multinomial choice model discussed in Section *III.C.2* , judges classify individuals into incarceration, conviction, and dismissal depending on the cases' unobservable V_i and $\tilde{V_c}$. Panel (a) visualizes this for an arbitrary judge, Panel (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.

Figure IV: Distribution of the stringency instruments

(a) Noncarceral conviction

(b) Incarceration

(c) Scatter plot of residualized instruments
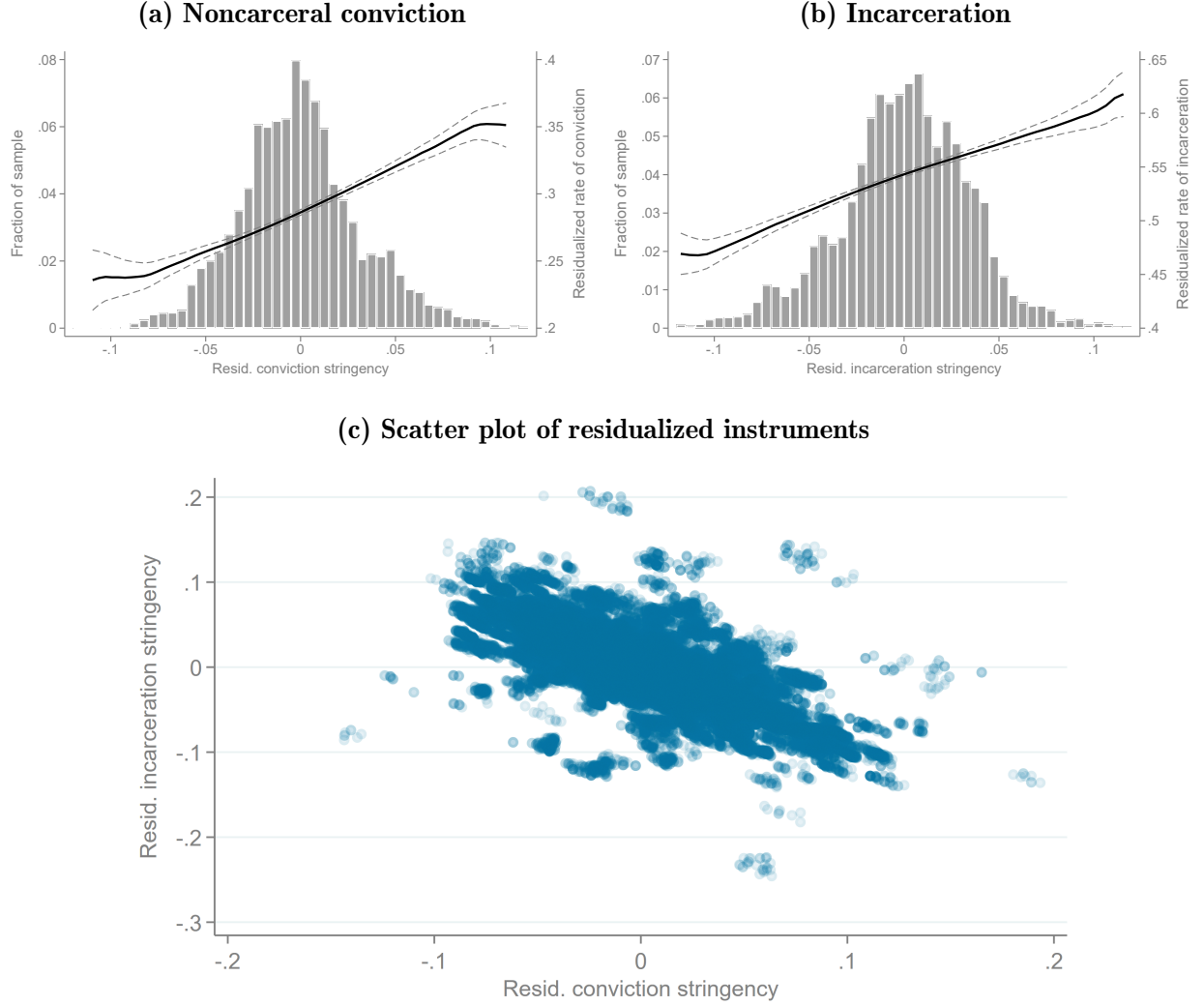
Note: This figure presents our first stages in graphical format for noncarceral conviction (Panel a) and incarceration (Panel b). The histograms plot the density of our residualized measures of conviction or incarceration stringency, and the line plots estimates of the first stage regression with conviction (Panel a) and incarceration (Panel b) as the dependent variable. Panel c is a scatter plot of the residualized incarceration and conviction instruments. In all three panels, the corresponding instrument is residualized against day-of-the-week, court-by-month, and court-by-year fixed effects.

ONLINE APPENDIX

Conviction, Incarceration, and Recidivism: Understanding the Revolving Door

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

Contents

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, Pecenco, and Weaver (2021) with an additional label for Virginia. Following Norris, Pecenco, and Weaver (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.[1]

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% three-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.[2] Virginia's prison incarceration rate, shown in Panel (a), is 447 per 100,000 people. This rate 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.[3] 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 (Figure A.4). 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 difference is because discretionary parole was virtually abolished in Virginia for felonies committed after 1995, with inmates being required to serve at least 85% of their sentences, with the possibility to earn good-time credits toward early release. The initial carceral sentence is more closely linked to time spent incarcerated than in other places.

---

[1]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/.

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

[3]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
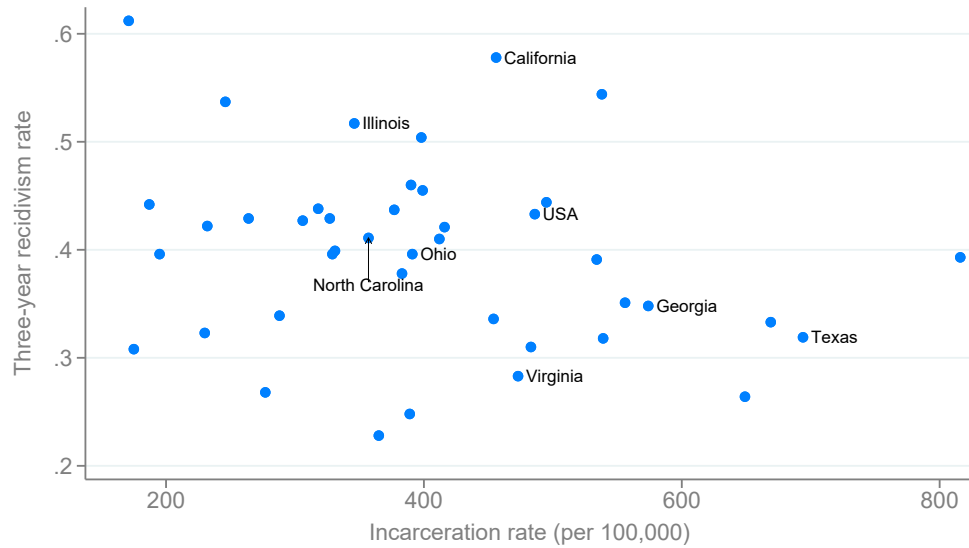
(b) Violent crime rates vs property crime rates
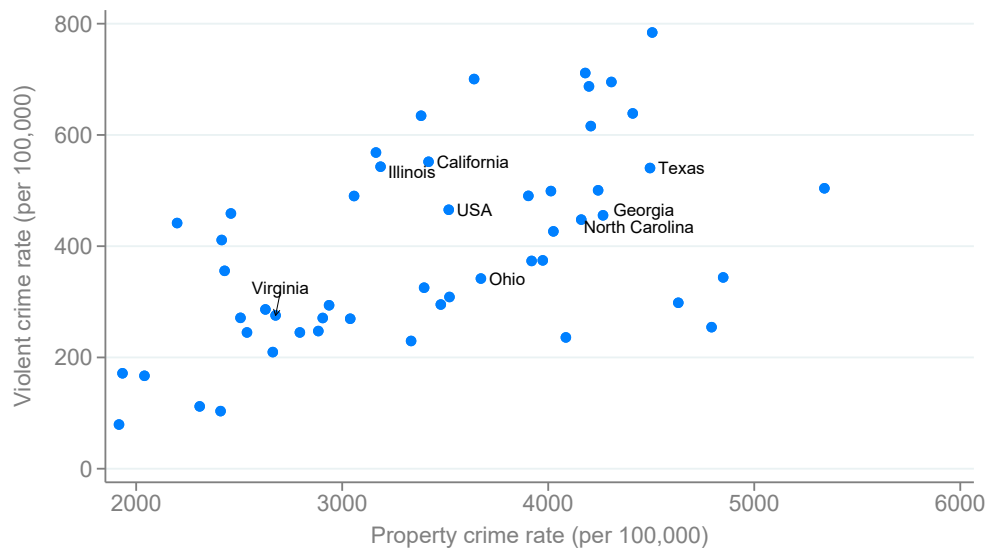
Note: Scatterplots of 2004 incarceration rates, 2004 three-year recidivism rates, and 2004 crime rates. Data gathered from the Pew Center, Bureau of Justice Statistics, and the FBI Uniform Crime Reporting Program.

Figure A.2: Incarceration rates

(a) Prison incarceration rates

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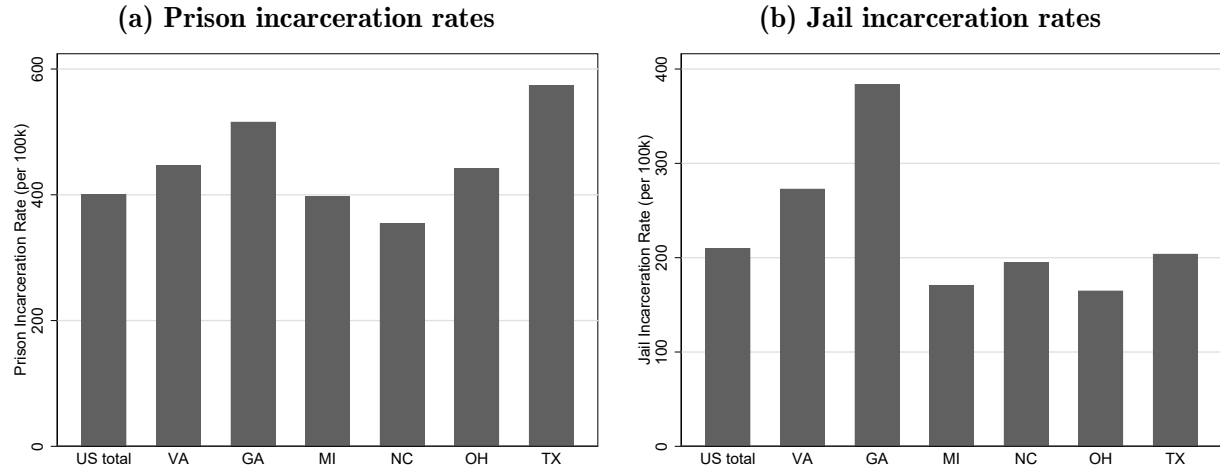
Note: This figure shows the prison (Panel A) and jail (Panel B) incarceration rates, respectively, per 100,000 residents for Virginia, the U.S. overall, Georgia, Michigan, North Carolina, Ohio, and Texas. Based on 2017 and 2014 data respectively from the Prison Policy Initiative (December 2018 press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

Figure A.3: Racial and ethnic composition of the imprisoned population
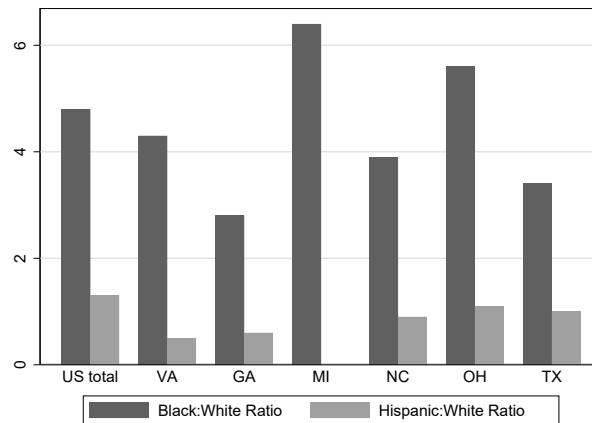
Note: This figure plots the ratio of incarceration rates for Black vs White residents (darker bars) and Hispanic vs White residents (lighter bars), for Virginia, the U.S. overall, Georgia, Michigan, North Carolina, Ohio, and Texas in 2019. Data from sentencingproject.org, used to calculate incarceration by ethnicity, is not available for Michigan.

Figure A.4: Supervision rates

(a) Probation rates
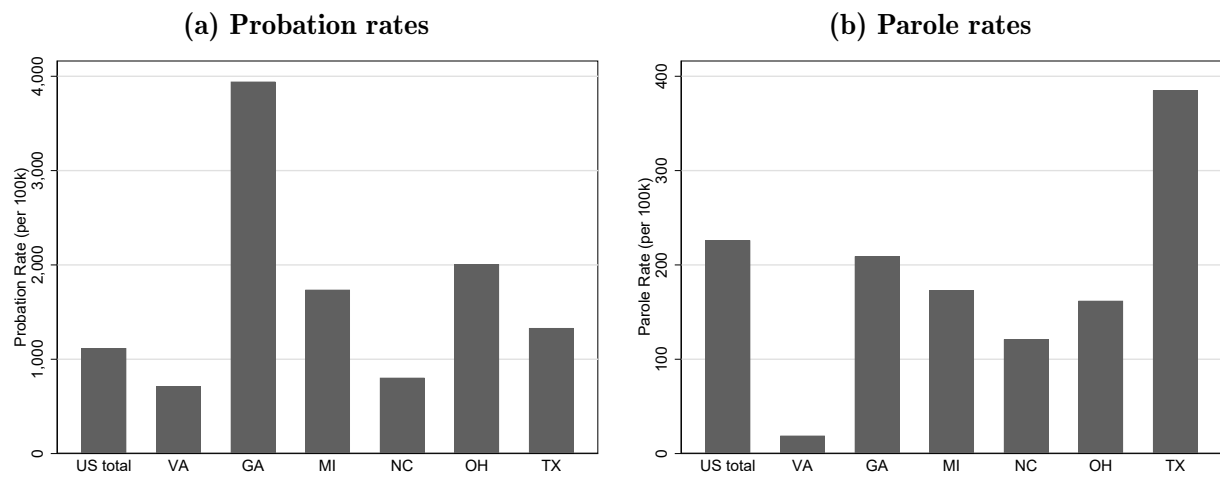
(b) Parole rates

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 source

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 Virginia's Circuit Courts (2021) 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. The VCC data constitute the primary data source for our 2SLS analysis with judge stringencies.

B.2   Supplementary data sources

Virginia Criminal Sentencing Commission (VCSC) data. The Virginia Criminal Sentencing Commission (2021) provided a dataset that contains information on individuals in Virginia sentenced for a felony. We use the VCSC data as a supplementary dataset for our 2SLS analysis (to construct our measure of prior convictions) and as the main source for the RD analysis. The VCSC data 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. It does not contain information on demographics or prior and future charges, so we match it to data from Virginia's Circuit Courts as described below.

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 Virginia's District Courts (2021) 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 that 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 IRS data (2005) produced by the IRS of average zip code earnings. We use the 2005 vintage and match in by zip onto our samples for supplementary analyses in 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. We also drop roughly 77,000 dockets (less than 5% of the remaining sample) that are missing disposition date or initiation date, as well as cases where the disposition is on a weekend.

- Sometimes prosecutors file separate dockets for different charges against the same defendant, for instance, if 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 after 2010 due to decreased data availability. Overall, we drop 121,931 cases, (27% of remaining cases), leaving 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 for various values of X. The VCC data goes through 2019. Our main sample includes the 183,381 cases disposed prior to 2012 to have seven years post-disposition for all cases. In a robustness check, we expand the sample to include cases disposed through 2015 when evaluating recidivism in years 2-4 and through 2018 when evaluating recidivism in year 1.

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 VCC 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 checks.

Incarceration-length RD data. The first sample leverages the discontinuity in the incarceration-length score as calculated in Worksheet A. We use that sample 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 prison sentence, since we do not 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, 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 is not smooth, potentially due to the scoring of worksheets for those categories. 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) and provides race, gender, arrest date, and prior charges. We drop cases from Fairfax and Alexandria, which are not in the VCC data. We use the fuzzy matching method developed by Enamorado, Fifield, and Imai (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 (269,437 cases, or 59% of the data).

- 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 VCC 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. Our sample size is then much smaller, 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 analyses come 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 I and G.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 carceral sentence.

- *Noncarceral 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, 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, this outcome is similar to the previous variable, 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. We have 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.[4]

- *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.

---

[4]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.

- *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.

- *% 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.

C PROOFS AND DERIVATIONS

C.1 Proof of Proposition 1

When CPM holds but UPM does not, a shift 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

$$E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))] = \omega_{d \to c} \Delta_{d \to c}^{Y_c - Y_d} + \omega_{i \to c} \Delta_{i \to c}^{Y_c - Y_i} + \omega_{d \to i} \Delta_{d \to i}^{Y_i - Y_d}. \tag{1}$$

Since the overall probability of incarceration is fixed at z_i, the share of cases flowing into and out of incarceration must be equal, $\omega_{d \to i} = \omega_{i \to c}$. Hence, we can rewrite appendix equation (1) as

$$E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))] = \omega_{d \to c} \Delta_{d \to c}^{Y_c - Y_d} + \omega_{i \to c} \left[\Delta_{i \to c}^{Y_c - Y_i} + \Delta_{d \to i}^{Y_i - Y_d} \right]. \tag{2}$$

Next, observe that

$$\Delta_{d \to i}^{Y_i - Y_d} = \Delta_{d \to i}^{Y_i - Y_c} + \Delta_{d \to i}^{Y_c - Y_d}.$$

Appendix equation (2) can be rewritten as

$$\begin{aligned} E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))] &= \omega_{d \to c} \Delta_{d \to c}^{Y_c - Y_d} + \omega_{i \to c} \left[\Delta_{i \to c}^{Y_c - Y_i} + \Delta_{d \to i}^{Y_i - Y_c} + \Delta_{d \to i}^{Y_c - Y_d} \right] \\ &= \omega_{d \to c} \Delta_{d \to c}^{Y_c - Y_d} + \omega_{i \to c} \Delta_{d \to i}^{Y_c - Y_d} + \omega_{i \to c} \left[\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c} \right]. \end{aligned}$$

For the denominator of the Wald estimand, we have

$$E[T_c(z_c', z_i) - T_c(z_c, z_i)] = \omega_{d \to c} + \omega_{i \to c}.$$

Constructing the Wald estimand, we obtain equation (9):

$$\frac{E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} =$$

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

Moving from CPM to the stronger UPM assumption simplifies equation (9). First, recall that $UPM(Z_c \mid Z_i)$ 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 stringency $Z_i = z_i$ 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 $UPM(Z_c \mid Z_i)$ rules out flows 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 (9) to

$$\frac{E[Y(T(z'_c, z_i)) - Y(T(z_c, z_i))]}{E[T_c(z'_c, z_i) - T_c(z_c, z_i)]} = \Delta_{d \to c}^{Y_c - Y_d}.$$

C.2 Bias with four treatments

Here, we calculate the asymptotic bias for 2SLS 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 III.A 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 represents 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) $l \to s$. This set 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 of $l \to s$; but under CPM we can only have one, not both. The same applies for (5).

As with 3 treatments, holding z_s fixed means that flows into and out of $T = s$ have to be equal, and holding z_l fixed means flows into 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

$$E[Y(T(z'_c, z_s, z_l)) - Y(T(z_c, z_s, z_l))] = \tag{3}$$
$$\Delta_{d \to c}^{Y_c - Y_d}\omega_{d \to c} + [\Delta_{d \to s}^{Y_s - Y_d}\omega_{d \to s} + \Delta_{s \to c}^{Y_c - Y_s}\omega_{s \to c}] + [\Delta_{d \to l}^{Y_l - Y_d}\omega_{d \to l} + \Delta_{l \to c}^{Y_c - Y_l}\omega_{l \to c}] + \Delta_{l \to s}^{Y_s - Y_l}\omega_{l \to s},$$

where brackets have been placed around two sets of terms to simplify our discussion.

For any difference in two potential outcomes, we can always rewrite it as $Y_k - Y_j = (Y_k - Y_m) - (Y_j - Y_m)$. The first term in the square brackets in appendix equation (3) can be rewritten as

$$[\Delta_{d \to s}^{Y_s - Y_d}\omega_{d \to s} + \Delta_{s \to c}^{Y_c - Y_s}\omega_{s \to c}] = [\Delta_{d \to s}^{Y_s - Y_d}\omega_{d \to s} + \Delta_{s \to c}^{Y_c - Y_s}(\omega_{d \to s} + \omega_{l \to s})] \tag{4}$$
$$= [\Delta_{d \to s}^{Y_s - Y_d}\omega_{d \to s} + (\Delta_{s \to c}^{Y_c - Y_d} - \Delta_{s \to c}^{Y_s - Y_d})\omega_{d \to s} + \Delta_{s \to c}^{Y_c - Y_s}\omega_{l \to s}]$$
$$= [\Delta_{s \to c}^{Y_c - Y_d}\omega_{d \to s} + (\Delta_{d \to s}^{Y_s - Y_d} - \Delta_{s \to c}^{Y_s - Y_d})\omega_{d \to s} + \Delta_{s \to c}^{Y_c - Y_s}\omega_{l \to s}].$$

The second term in the square brackets from appendix equation (3) can be rewritten as

$$[\Delta_{d \to l}^{Y_l - Y_d}\omega_{d \to l} + \Delta_{l \to c}^{Y_c - Y_l}\omega_{l \to c}] = [\Delta_{d \to l}^{Y_l - Y_d}(\omega_{l \to s} + \omega_{l \to c}) + \Delta_{l \to c}^{Y_c - Y_l}\omega_{l \to c}] \tag{5}$$
$$= [\Delta_{d \to l}^{Y_l - Y_d}\omega_{l \to s} + \Delta_{d \to l}^{Y_l - Y_d}\omega_{l \to c} + (\Delta_{l \to c}^{Y_c - Y_d} - \Delta_{l \to c}^{Y_l - Y_d})\omega_{l \to c}]$$
$$= [\Delta_{d \to l}^{Y_l - Y_d}\omega_{l \to s} + \Delta_{l \to c}^{Y_c - Y_d}\omega_{l \to c} + (\Delta_{d \to l}^{Y_l - Y_d} - \Delta_{l \to c}^{Y_l - Y_d})\omega_{l \to c}].$$

So, appendix equation (3) can be written as

$$E[Y(T(z_c', z_s, z_l)) - Y(T(z_c, z_s, z_l))] = \tag{6}$$

$$\Delta_{d\to c}^{Y_c-Y_d}\omega_{d\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{d\to s} + \Delta_{l\to c}^{Y_c-Y_d}\omega_{l\to c}$$
$$+ (\Delta_{d\to s}^{Y_s-Y_d} - \Delta_{s\to c}^{Y_s-Y_d})\omega_{d\to s}$$
$$+ (\Delta_{d\to l}^{Y_l-Y_d} - \Delta_{l\to c}^{Y_l-Y_d})\omega_{l\to c}$$
$$+ \Delta_{l\to s}^{Y_s-Y_l}\omega_{l\to s} + \Delta_{s\to c}^{Y_c-Y_s}\omega_{l\to s} + \Delta_{d\to l}^{Y_l-Y_d}\omega_{l\to s},$$

and the last row of appendix equation (6) can be rewritten as

$$\Delta_{l\to s}^{Y_s-Y_l}\omega_{l\to s} + \Delta_{s\to c}^{Y_c-Y_s}\omega_{l\to s} + \Delta_{d\to l}^{Y_l-Y_d}\omega_{l\to s} \tag{7}$$
$$= \Delta_{l\to s}^{Y_s-Y_l}\omega_{l\to s} + (\Delta_{s\to c}^{Y_c-Y_d}\omega_{l\to s} - \Delta_{s\to c}^{Y_s-Y_d}\omega_{l\to s}) + \Delta_{d\to l}^{Y_l-Y_d}\omega_{l\to s}$$
$$= \Delta_{l\to s}^{Y_s-Y_l}\omega_{l\to s} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{l\to s} - (\Delta_{s\to c}^{Y_s-Y_l}\omega_{l\to s} + \Delta_{s\to c}^{Y_l-Y_d}\omega_{l\to s}) + \Delta_{d\to l}^{Y_l-Y_d}\omega_{l\to s}$$
$$= \Delta_{s\to c}^{Y_c-Y_d}\omega_{l\to s} + (\Delta_{l\to s}^{Y_s-Y_l} - \Delta_{s\to c}^{Y_s-Y_l})\omega_{l\to s} + (\Delta_{d\to l}^{Y_l-Y_d} - \Delta_{s\to c}^{Y_l-Y_d})\omega_{l\to s}.$$

Rewriting appendix equation (6), we get

$$E[Y(T(z_c', z_s, z_l)) - Y(T(z_c, z_s, z_l))] = \tag{8}$$

$$\Delta_{d\to c}^{Y_c-Y_d}\omega_{d\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{d\to s} + \Delta_{l\to c}^{Y_c-Y_d}\omega_{l\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{l\to s}$$
$$+ (\Delta_{d\to s}^{Y_s-Y_d} - \Delta_{s\to c}^{Y_s-Y_d})\omega_{d\to s}$$
$$+ (\Delta_{d\to l}^{Y_l-Y_d} - \Delta_{l\to c}^{Y_l-Y_d})\omega_{l\to c}$$
$$+ (\Delta_{l\to s}^{Y_s-Y_l} - \Delta_{s\to c}^{Y_s-Y_l})\omega_{l\to s} + (\Delta_{d\to l}^{Y_l-Y_d} - \Delta_{s\to c}^{Y_l-Y_d})\omega_{l\to s}.$$

The next step is to rewrite the first row of appendix equation (8) as

$$\Delta_{d\to c}^{Y_c-Y_d}\omega_{d\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{d\to s} + \Delta_{l\to c}^{Y_c-Y_d}\omega_{l\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{l\to s} \tag{9}$$
$$= \Delta_{d\to c}^{Y_c-Y_d}\omega_{d\to c} + \Delta_{s\to c}^{Y_c-Y_d}(\omega_{d\to s} + \omega_{l\to s}) + \Delta_{l\to c}^{Y_c-Y_d}\omega_{l\to c}$$
$$= \Delta_{d\to c}^{Y_c-Y_d}\omega_{d\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{s\to c} + \Delta_{l\to c}^{Y_c-Y_d}\omega_{l\to c}.$$

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

$$E[Y(T(z_c', z_s, z_l)) - Y(T(z_c, z_s, z_l))] = \tag{10}$$
$$\underbrace{\Delta_{d\to c}^{Y_c-Y_d}\omega_{d\to c} + \Delta_{s\to c}^{Y_c-Y_d}\omega_{s\to c} + \Delta_{l\to c}^{Y_c-Y_d}\omega_{l\to c}}_{\text{Weighted } d \to c \text{ treatment effects}}$$

$$+ (\Delta_{d\to s}^{Y_s-Y_d} - \Delta_{s\to c}^{Y_s-Y_d})\omega_{d\to s}$$
$$+ (\Delta_{d\to l}^{Y_l-Y_d} - \Delta_{l\to c}^{Y_l-Y_d})\omega_{l\to c}$$
$$+ (\Delta_{l\to s}^{Y_s-Y_l} - \Delta_{s\to c}^{Y_s-Y_l})\omega_{l\to s}$$
$$\underbrace{+ (\Delta_{d\to l}^{Y_l-Y_d} - \Delta_{s\to c}^{Y_l-Y_d})\omega_{l\to s}}_{\text{Differences in subgroup treatment effects}} \quad .$$

We can similarly express the denominator of the Wald estimator as

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

Finally, dividing appendix equation (10) by appendix equation (11), we end up with two terms. The first term is a weighted average of margin-specific treatment effects when moving from $T = d$ to $T = c$. The weights here are taken from three groups of compliers, are 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.[5] The bias depends on how heterogeneous the treatment effects are. For example, under a constant effects assumption, the bias terms are all zero.

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, and the size of the asymptotic 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 what changes when swapping out CPM for the unordered partial monotonicity (UPM) assumption. Specifically, we consider the case where we 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

$$
\begin{aligned}
\text{Wald}(z_c', z_c \mid z_i) &\equiv \frac{E[Y(T(z_c', z_i)) - Y(T(z_c, z_i))]}{E[T_c(z_c', z_i) - T_c(z_c, z_i)]} \\
&= \frac{E[Y \mid Z_c = z_c', Z_i = z_i] - E[Y \mid Z_c = z_c, Z_i = z_i]}{E[T_c \mid Z_c = z_c', Z_i = z_i] - E[T_c \mid Z_c = z_c, Z_i = z_i]} \\
&= \underbrace{\frac{\omega_{d \to c} \Delta_{d \to c}^{Y_c - Y_d} + \omega_{i \to c} \Delta_{d \to i}^{Y_c - Y_d}}{\omega_{d \to c} + \omega_{i \to c}}}_{\text{Weighted avg. of } Y_c - Y_d \text{ treatment effects}} + \underbrace{\frac{\omega_{i \to c}}{\omega_{d \to c} + \omega_{i \to c}} \left[\Delta_{d \to i}^{Y_i - Y_c} - \Delta_{i \to c}^{Y_i - Y_c} \right]}_{\text{Bias term}}.
\end{aligned}
$$

Now, we derive what is identified by IV when using multiple judges with varying conviction stringency but the same incarceration stringency. For readability, we leave the conditioning

---

[5]As discussed above, $\omega_{s \to c} = \omega_{l \to s} + \omega_{d \to s}$ and $\omega_{d \to l} = \omega_{l \to s} + \omega_{l \to c}$. With these two identities, it is straightforward to show that $\omega_{d \to c} + \omega_{s \to c} + \omega_{l \to c} = \omega_{d \to s} + \omega_{l \to c} + \omega_{l \to s} + \omega_{l \to s}$, making the second term a weighted average of the four bias terms.

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), first consider the numerator:

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

Next, we can rewrite the denominator as

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

Putting these together, we get

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

where

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

Other than the implicit conditioning on $Z_i = z_i$, this formula is the same as the formula derived in Imbens and Angrist (1994). However, in our setting, the Wald estimand may not always be a pairwise LATE as in Imbens and Angrist (1994). Under the CPM assumption and A1–A4, the Wald estimand recovers the term given in equation (9) in Section III. 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 (9) simplifies 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 settings (Angrist and Pischke, 2009; Blandhol et al., 2022).

Table C.1 shows estimates from a specification where the treatment and instrument have been interacted with the other stringency for that judge. Specification 1 interacts the instrument and treatment with residualized terciles of the other stringency and includes our standard set of fixed effects: court-by-year, court-by-month, and day-of-week dummies. Specification 2 replaces court-by-year and court-by-month dummies with court-by-year-by-month dummies. Since the tercile interactions only condition on three bins of incarceration or dismissal stringency, Specification 3 further adds dummies for deciles of residualized judge incarceration or dismissal stringency. Specification 4 replaces the conviction instrument or incarceration instrument interacted with residualized incarceration or dismissal stringency terciles with judge dummies. Some caution should be taken in interpreting these estimates, as splitting our sample into thirds quickly leads to large standard errors and small first-stage F-statistics. We report four specifications that include increasingly rich sets of controls. For conviction, most of the estimates are positive, nearly all estimates are positive when including richer controls, and all negative estimates are statistically insignificant with very large standard errors. We see similar trends with large impacts of conviction in the first year that accumulate over time.

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

	Impacts of conviction with incarceration stringency bins				Impacts of incarceration with dismissal stringency bins			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Specification 1								
Convict x bottom 3rd	0.117	-0.046	0.119	0.158	-0.018	-0.108	0.039	-0.115
	(0.075)	(0.140)	(0.108)	(0.166)	(0.066)	(0.088)	(0.089)	(0.117)
Convict x middle 3rd	0.267**	0.192	0.289	0.584**	-0.180**	0.023	-0.050	-0.075
	(0.125)	(0.282)	(0.180)	(0.292)	(0.073)	(0.095)	(0.091)	(0.133)
Convict x top 3rd	0.320	-1.057	0.470	-0.124	-0.029	0.017	0.055	0.089
	(0.437)	(0.967)	(0.644)	(0.959)	(0.054)	(0.093)	(0.075)	(0.114)
Specification 2								
Convict x bottom 3rd	0.107	-0.038	0.159	0.205	-0.019	-0.105	0.014	-0.139
	(0.089)	(0.148)	(0.129)	(0.196)	(0.071)	(0.093)	(0.094)	(0.124)
Convict x middle 3rd	0.278*	0.123	0.372	0.640*	-0.187**	0.004	-0.039	-0.094
	(0.155)	(0.306)	(0.229)	(0.361)	(0.077)	(0.104)	(0.096)	(0.147)
Convict x top 3rd	0.201	-0.973	0.466	-0.065	-0.030	0.056	0.067	0.124
	(0.394)	(0.752)	(0.601)	(0.873)	(0.057)	(0.096)	(0.079)	(0.119)
Specification 3								
Convict x bottom 3rd	0.071	0.115	0.132	0.277**	-0.072	-0.098	0.026	-0.146
	(0.055)	(0.086)	(0.088)	(0.122)	(0.073)	(0.081)	(0.080)	(0.109)
Convict x middle 3rd	0.111	0.128	0.218*	0.397**	-0.256	-0.008	-0.046	-0.146
	(0.075)	(0.122)	(0.117)	(0.165)	(0.208)	(0.208)	(0.204)	(0.291)
Convict x top 3rd	0.109	-0.060	0.237	0.269	-0.050	0.061	-0.030	0.047
	(0.102)	(0.148)	(0.156)	(0.228)	(0.135)	(0.176)	(0.155)	(0.231)
Specification 4								
Convict x bottom 3rd	0.032	-0.006	0.008	0.031	-0.104***	-0.070*	-0.057*	-0.142***
	(0.028)	(0.043)	(0.038)	(0.054)	(0.023)	(0.036)	(0.032)	(0.044)
Convict x middle 3rd	0.021	0.023	0.008	0.054	-0.066**	-0.044	-0.044	-0.095*
	(0.034)	(0.051)	(0.048)	(0.068)	(0.027)	(0.041)	(0.036)	(0.052)
Convict x top 3rd	0.060**	0.034	0.084**	0.101*	-0.004	0.078*	0.069*	0.137***
	(0.030)	(0.049)	(0.042)	(0.061)	(0.025)	(0.040)	(0.036)	(0.052)
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table instruments for three endogenous variables in each 2SLS regression: conviction interacted with residualized incarceration stringency terciles (Columns 1-4) or incarceration interacted with residualized dismissal stringency terciles (Columns 5-8). In Specification 1, the instruments in each first stage are conviction stringency interacted with residualized incarceration stringency terciles (Columns 1-4) or incarceration stringency interacted with residualized dismissal stringency terciles (Columns 5-8). The regression includes our standard fixed effects: court-by-year, court-by-month, and day-of-week dummies. Specification 2 is the same as Specification 1, except we replace court-by-year and court-by-month dummies with court-by-year-by-month dummies. Specification 3 is the same as Specification 2, but it adds dummies for deciles of residualized judge incarceration (Columns 1-4) or dismissal stringency (Columns 5-8). Specification 4 is the same as Specification 3, except it instead uses only judge dummies as instruments. Standard errors are clustered at the judge-year level. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01.

C.4   Average UPM

$UPM(Z_c \mid Z_i)$ 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, Lefgren, and Leslie

(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, Lefgren, and Leslie's (2023) average monotonicity condition into the three treatment setting and refer to this as "average UPM$(Z_c \mid Z_i)$." We focus on the condition that is relevant to the specification where we are instrumenting for conviction and controlling for incarceration stringency; average UPM$(Z_i \mid Z_d)$ 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 UPM($Z_c \mid Z_i$).

For all (g, z_i) in the support of (G, Z_i) the following conditions must hold:

$$Cov(T_c(Z_c, Z_i), Z_c \mid Z_i = z_i, G = g) \geq 0$$
$$Cov(T_i(Z_c, Z_i), Z_c \mid Z_i = z_i, G = g) = 0.$$

To illustrate a difference between UPM$(Z_c \mid Z_i)$ and average UPM$(Z_c \mid Z_i)$, 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 move them from conviction to dismissal, UPM$(Z_c \mid Z_i)$ would be violated but average UPM$(Z_c \mid Z_i)$ might not be. As long as the probability of conviction for each group is positively correlated with the overall conviction propensity of judges, average UPM$(Z_c \mid Z_i)$ is satisfied.

Average UPM$(Z_c \mid Z_i)$, along with A1–A4, 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) ensure that the exogeneity condition outlined in Blandhol et al. (2022) is met. In our setting, this exogeneity condition means that $G, Y(T = c) \perp Z_c \mid Z_i$. 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 covariates 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 a 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.[6] They

---

[6]The necessary and sufficient condition for weakly causal estimates is presented in the paragraph between

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 occur naturally. For instance, a judge who is generally harsh may be relatively lenient on certain types of offenders, which 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 IV control for Z_i rather than conditioning on it. 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 III, as long as one additional assumption is met.

Blandhol et al. (2022) consider what 2SLS recovers when covariates are included as controls, but are 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 sub-group. They go 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.[7]

A4b. Rich covariates. The linear projection of Z on X is equal to the conditional expectation of Z given X. That is $L[Z \mid X] = X'E[XX']^{-1}E[XZ] = E[Z \mid X]$.

Assumption A4b implies that we need to include a rich set of controls. Note that assumption A4b differs from assumption A4 because Section III.B 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, 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.'s (2022) 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–A5 or A1–A4 and A6, these Wald estimands are equivalent to those we derive in Section III.D. Thus, under assumptions

---

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 $\mathbb{1}(Z \in \mathbb{Z}_j(g))$.

[7]Note that assumptions A1–A3, and A5 satisfy the other needed assumptions in Blandhol et al. (2022). In particular, A5 implies their "Ordered strong monotonicity" (OSM). Assumption A6 also satisfies the OSM, but violates their definition of exclusion, which can result in biased Wald estimates, similar to those we derive under CPM.

A1–A3, A4b and A6, 2SLS recovers a positively-weighted average of terms that are margin-specific causal effects plus additive bias terms. Under assumptions A1–A3, A4b, and A5, 2SLS recovers a positively-weighted average of margin-specific treatment effects.

Table C.2 shows 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 judge's other 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 a 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 years, or one to seven years.

Table C.2: The impacts of conviction and incarceration on recidivism: robustness to alternative specifications and controls

	Impacts of conviction				Impacts of incarceration			
	(1) Year 1	(2) Years 2-4	(3) Years 5-7	(4) Years 1-7	(5) Year 1	(6) Years 2-4	(7) Years 5-7	(8) Years 1-7
Specification 1								
Fut. charge	0.101**	0.126	0.104	0.292***	-0.101***	-0.048	-0.010	-0.099*
	(0.051)	(0.083)	(0.080)	(0.109)	(0.029)	(0.046)	(0.041)	(0.059)
Fut. conviction	0.134***	0.163**	0.074	0.357***	-0.113***	-0.068	0.006	-0.134**
	(0.048)	(0.079)	(0.076)	(0.106)	(0.028)	(0.046)	(0.039)	(0.058)
Fut. incarceration	0.103**	0.088	-0.005	0.256***	-0.076***	-0.022	0.036	-0.061
	(0.042)	(0.069)	(0.061)	(0.093)	(0.024)	(0.039)	(0.032)	(0.050)
Specification 2								
Fut. charge	0.089	0.156*	0.154	0.354***	-0.103***	-0.042	-0.017	-0.107*
	(0.060)	(0.094)	(0.095)	(0.128)	(0.031)	(0.050)	(0.044)	(0.064)
Fut. conviction	0.128**	0.193**	0.123	0.430***	-0.118***	-0.065	0.003	-0.142**
	(0.057)	(0.090)	(0.091)	(0.126)	(0.030)	(0.049)	(0.042)	(0.062)
Fut. incarceration	0.107**	0.111	0.035	0.332***	-0.079***	-0.027	0.031	-0.075
	(0.048)	(0.079)	(0.074)	(0.110)	(0.026)	(0.042)	(0.034)	(0.054)
Specification 3								
Fut. charge	0.116**	0.157*	0.098	0.324***	-0.099***	-0.039	-0.014	-0.093
	(0.054)	(0.089)	(0.081)	(0.115)	(0.031)	(0.047)	(0.042)	(0.062)
Fut. conviction	0.153***	0.197**	0.071	0.407***	-0.116***	-0.060	-0.007	-0.131**
	(0.053)	(0.085)	(0.077)	(0.113)	(0.030)	(0.046)	(0.040)	(0.060)
Fut. incarceration	0.121***	0.124*	-0.021	0.298***	-0.075***	-0.020	0.024	-0.069
	(0.045)	(0.074)	(0.063)	(0.099)	(0.025)	(0.039)	(0.033)	(0.053)
Specification 4								
Fut. charge	0.108	0.197*	0.149	0.396***	-0.102***	-0.044	-0.034	-0.119*
	(0.067)	(0.103)	(0.099)	(0.140)	(0.034)	(0.051)	(0.045)	(0.066)
Fut. conviction	0.154**	0.231**	0.120	0.493***	-0.120***	-0.068	-0.024	-0.160**
	(0.065)	(0.100)	(0.095)	(0.141)	(0.033)	(0.050)	(0.043)	(0.064)
Fut. incarceration	0.131**	0.151*	0.013	0.384***	-0.077***	-0.038	0.010	-0.103*
	(0.054)	(0.086)	(0.077)	(0.123)	(0.028)	(0.043)	(0.036)	(0.057)
Observations	183,371	183,371	183,371	183,371	183,371	183,371	183,371	183,371

Note: This table reports estimates of the impact of conviction and incarceration on our three measures of recidivism. Specification 1 includes the fixed effects from the main specification (court-by-year, court-by-month, and day of week dummies) plus 100 percentile dummies for residualized judge incarceration stringencies (first four columns) and dismissal stringencies (last four columns). Specification 2 matches specification 1 but swaps out court-by-year and court-by-month fixed effects with court-by-year-by-month fixed effects. Specification 3 includes the fixed effects from the main regression plus fixed effects for the year-by-decile of residualized incarceration stringencies (first four columns) and dismissal stringencies (last four columns). Specification 4 is the same as specification 3, but swaps out court-by-year and court-by-month dummies with court-by-year-by-month dummies. Standard errors are clustered at the judge-year level. Stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.6 2SLS with two endogenous variables

Here we briefly discuss why the specifications in (1)-(2) and (3)-(4) should have the same estimand as what would be obtained using a single 2SLS regression with two endogenous treatment variables and both stringencies.

In the population, we should have $\gamma_0 = 0$, $\gamma_1 = 1$, and $\gamma_2 = 0$ for the coefficients in equation (3). Thus, δ_1 in (4) should be equal to δ_1' in the following regression:

$$Y = \delta_0' + \delta_1' Z_c + \delta_2' Z_i + \nu'.$$

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

$$T_c = \gamma_0 + \gamma_1 Z_c + \gamma_2 Z_i + \epsilon$$
$$T_i = \omega_0 + \omega_1 Z_c + \omega_2 Z_i + \upsilon$$
$$Y = \delta_0'' + \delta_1'' T_c + \delta_2'' T_i + \nu''.$$

By similar logic, $\omega_0 = 0$, $\omega_1 = 0$, and $\omega_2 = 1$. Thus, $\delta_1 = \delta_1' = \delta_1''$ and $\delta_2 = \delta_2' = \delta_2''$.

In our sample, the first-stage coefficients are not precisely zero or one, as is common in finite samples. Yet, these two approaches produce similar estimates. Table C.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 $T_{\backslash d}$ (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 $T = d$.

C.7 2SLS with a binary treatment indicator

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

$$T_i = \gamma_0 + \gamma_1 Z_i + \epsilon$$
$$Y = \delta_0 + \delta_1 T_i + \nu.$$

This specification is similar to the specification in equations (1) and (2) from the main text, but does not include judge dismissal stringency as a control. Under the standard LATE assumptions, δ_1 will not represent 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 since $Z_i = 1 - (Z_c + Z_d)$ by construction.

Table C.1: Two instruments and two endogenous variables

	Year 1		Years 2-4		Years 5-7		Years 1-7	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Convict: fut. charge	-0.002	0.105**	0.004	0.091	0.006**	0.083	0.011***	0.244**
	(0.002)	(0.048)	(0.003)	(0.078)	(0.002)	(0.078)	(0.004)	(0.100)
Incar: fut. charge	-0.022***	-0.097***	0.013***	-0.017	0.025***	0.004	0.022***	-0.071
	(0.002)	(0.029)	(0.002)	(0.047)	(0.002)	(0.040)	(0.003)	(0.059)
Convict: fut. conv.	0.001	0.136***	0.008***	0.120	0.007***	0.060	0.014***	0.311***
	(0.002)	(0.044)	(0.003)	(0.075)	(0.002)	(0.074)	(0.004)	(0.098)
Incar: fut. conv.	-0.018***	-0.112***	0.013***	-0.038	0.023***	0.021	0.022***	-0.107*
	(0.001)	(0.029)	(0.002)	(0.047)	(0.002)	(0.038)	(0.003)	(0.058)
Convict: fut. incar.	0.001	0.115***	0.006**	0.063	0.005**	-0.023	0.012***	0.226***
	(0.002)	(0.039)	(0.002)	(0.066)	(0.002)	(0.059)	(0.003)	(0.086)
Incar: fut. incar.	-0.010***	-0.071***	0.017***	0.007	0.021***	0.053	0.027***	-0.030
	(0.001)	(0.024)	(0.002)	(0.041)	(0.002)	(0.032)	(0.003)	(0.051)
Ctrl Mean: fut. charge	0.088	0.088	0.175	0.175	0.132	0.132	0.306	0.306
Ctrl Mean: fut. conv.	0.077	0.077	0.159	0.159	0.120	0.120	0.283	0.283
Ctrl Mean: fut. incar.	0.055	0.055	0.115	0.115	0.084	0.084	0.212	0.212
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table shows IV and OLS estimates of the impacts of conviction and incarceration on future recidivism, where the IV results instrument for both endogenous variables within the same regression. The first six rows report the estimated impact of conviction or incarceration on different measures of recidivism: any future charge, any future conviction, and any future incarceration. Recidivism is measured from the time of sentencing and within the time window shown at top. For the OLS estimates, we regress our measures of recidivism on a dummy for conviction (regardless of incarceration status) and a dummy for incarceration. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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.

D   VALIDATING ASSUMPTIONS A1–A4

In this section, we discuss whether Assumptions A1–A4 from Section III 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 judges can influence both conviction and incarceration outcomes, expanding on Section *II.B*. We also present empirical evidence that judges influence 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 (trials by judge, without lay jurors). They also exert indirect influence on the likelihood of conviction through multiple channels. First, they make determinations 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. Moreover, their reputation as a tough or lenient judge might shape what offers prosecutors and defense attorneys are willing to put forward (LaCasse and Payne, 1999). For example, if the judge has a reputation for choosing short sentences, the prosecutor may adjust and offer shorter sentences as part of a plea deal.

We find persistent differences in case outcomes across judges. Panels A and B of Figure IV in the main paper shows the histogram of judge noncarceral conviction stringency (Panel A) and judge incarceration stringency (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.[8] Both panels plot the local linear regression of the residualized court outcome on the instrument.

Panel C of Figure IV plots the residualized noncarceral 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 II 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 noncarceral 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 165 and 360. Columns 4 through 6 perform similar first-stage regressions on 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 288 and 351.

Random assignment. As discussed in Section II.A, in our sample, cases are quasi-randomly assigned to judges within each court. There is either actual randomization, or case assignment is done based on scheduling or judge availability.[9] In addition, we confirm empirically that judge stringency is not predicted by case characteristics. In Table III of the main paper, we show that case characteristics are strong predictors of being convicted and of being incarcerated (columns 1 and 3). We then show that case characteristics largely do not predict judge conviction stringency (column 2) or incarceration stringency (column 4). For the few instances where covariates have statistically significant loadings, the predicted difference in stringency tends to be very small. Table D.1 replicates columns (2) and (4) from Table III but using standardized stringency measures. The odd columns regress non-carceral conviction stringency and incarceration stringency on case characteristics, where the stringency measure

---

[8]We constructed conviction stringency by residualizing an indicator for noncarceral conviction against county-by-year, county-by-month, and day-of-week fixed effects, then constructing leave-one-out averages at the judge-by-three-year level. We use a similar procedure to construct ncarceration stringency.

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

has been standardized to have a mean of zero and a standard deviation of one. The largest loading is on an indicator for assault cases, which predicts assault cases are associated with a 0.015 standard deviation change in stringency. The odd columns do not account for variation in stringency caused by variation over time or across courts. The even columns replicate the regressions from the odd columns but first residualize the stringency instruments for the set of circuit and time fixed effects. The largest coefficient in these regressions is 0.036. These results suggest that, while there are a few instances where covariates have statistically significant loadings, these loadings imply small predicted differences in stringency.

We show that our results are not sensitive to fully excluding certain types of cases from our analysis, either from the construction of the stringency instruments or from the 2SLS regressions. Table D.2 provides our main OLS and IV estimates for noncarceral conviction for four different subsets of cases. In Panel A we drop all cases involving assault charges when constructing the instrument and running the analyses, as assault is the offense type that is most predictive of both noncarceral conviction and incarceration stringency in our balance tables. These results are broadly similar to our main estimates in Table IV, with the same sign and magnitude. The two main differences are that the point estimates are moderately smaller, and standard errors are somewhat larger (likely due to the 15% reduction in sample size).

Panel B and C repeat the prior exercise, but throw out cases with drug offenses and cases with violent offenses, respectively. We focus on drug offenses and violent offenses since these are offense types where we believe judges may be most likely to differ in opinion on appropriate case outcomes. Dropping these offense types leads to broadly similar results, with similar point estimates and somewhat larger standard errors. Finally, Panel D drops cases with assault, sexual assault, fraud, or traffic charges (all offense types where there is any evidence of imbalance in Table III). Our estimates are again broadly similar. For this specification, we lose statistical significance on several coefficients that are significant in our main table, which may be in part due to moderately smaller (though similar in magnitude) estimates in Years 1–7, but is largely driven by larger standard errors, likely because of the 39% reduction in sample size. Table D.3 replicates the analysis in Table D.2, but for incarceration. These results follow the same pattern as the results for conviction. Overall, Tables D.2 and D.3 suggest that our results are not driven by potential exclusion violations.

In our robustness analysis in Appendix Section E, we compare how our estimates vary under several different assumptions. In Figures E.3–E.6, we show results where we use the full sample, but allow judge stringency to differ by (1) if the case has an assault charge or not and (2) if the case has a drug charge or not. These alternative constructions of our instrument are more demanding on our data, but we still find statistically significant increases in recidivism from non-carceral conviction 1–7 years after the case, and statistically significant decreases from incarceration only in the first year after the case.

Finally, as additional evidence of exogeneity, our first-stage estimates barely change when we add controls to our first-stage regression, as seen by comparing columns 2 and 3 and columns 5 and 6 of Table II in the main paper.

Exclusion. Our identification strategy relies on the assumption that the conviction stringency instrument only affects recidivism outcomes 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, a high-conviction judge giving longer sentences (holding incarceration probability fixed) 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.4, we find no evidence of a violation of the exclusion restriction for conviction. When we re-estimate the main IV regressions with an additional control for sentence length stringency or probability of sentence length shorter than 6 months and longer than 1 year and 4 years, our main conclusions do not change (see Appendix Figures E.3-E.6).[10]

A judge may influence other aspects of the case, such as probation and parole terms, or fines and fees. While we do not rule these channels out, we do not expect them to be as important. There are multiple 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., 2024; Lieberman, Luh, and Mueller-Smith, 2023). The findings in this literature bolster our confidence that, even if judge stringency in conviction and incarceration were correlated with other factors, they would not bias our results.

We do not expect decisions made at the beginning of the case, such as bail or pretrial detention, to lead to an exclusion violation. These decisions are made by bail magistrates that have no later influence over the case. Furthermore, there is often a month between the date of the arrest and when the defendant arrives at circuit court and the judge is assigned. It follows that 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 through 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.

Lastly, in Appendix Table E.1, we present reduced-form estimates, which regress outcomes on our instruments, and do not require the exclusion assumption to hold.

Monotonicity. One consequence of CPM (and the stronger condition, UPM) is that there will only be one-way flows across any margin. We present some empirical evidence to support this assumption. Following common practice for binary treatments (see, for example, Bhuller et al. (2020) or Norris, Pecenco, and Weaver (2021)), we conduct split-sample regressions where the data is bifurcated using observed characteristics such as race and gender. We then estimate judge stringency 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. Similarly, we would expect positive coefficients on all subsamples if Average UPM (defined in Appendix Subsection C.4) holds. Appendix Table D.5 reports the coefficient on the instrument from split-sample first-stage regressions. Each row presents a particular case characteristic. For

---

[10]We define sentence length stringency as the three-year 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.

example, the first row breaks our sample into whether a person has a drug charge or 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—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. Also see Section *IV.E* where we present a test of the UPM assumption.

Table D.1: Balance: outcomes in standard deviations

	Conv. string.	Resid. conv. string.	Incar. string.	Resid. incar. string.
	(1)	(2)	(3)	(4)
Any prior conv.	-0.0006	-0.0014	0.0030	0.0069
	(0.0026)	(0.0062)	(0.0027)	(0.0063)
Female	-0.0043*	-0.0103*	0.0025	0.0057
	(0.0023)	(0.0056)	(0.0024)	(0.0057)
Black	0.0032	0.0076	-0.0028	-0.0065
	(0.0022)	(0.0054)	(0.0022)	(0.0053)
Has misdemeanor	0.0013	0.0031	0.0041	0.0097
	(0.0037)	(0.0089)	(0.0039)	(0.0090)
Drugs	0.0045	0.0108	-0.0003	-0.0006
	(0.0032)	(0.0077)	(0.0035)	(0.0081)
Larceny	0.0035	0.0085	0.0041	0.0097
	(0.0027)	(0.0066)	(0.0029)	(0.0068)
Assault	-0.0148***	-0.0355***	0.0142***	0.0332***
	(0.0031)	(0.0075)	(0.0031)	(0.0072)
Fraud	0.0048	0.0114	0.0068*	0.0160*
	(0.0034)	(0.0082)	(0.0039)	(0.0090)
Traffic	-0.0036	-0.0087	0.0076*	0.0177*
	(0.0043)	(0.0103)	(0.0045)	(0.0104)
Burglary	-0.0016	-0.0039	0.0056	0.0132
	(0.0039)	(0.0093)	(0.0042)	(0.0098)
Robbery	-0.0026	-0.0062	0.0043	0.0101
	(0.0052)	(0.0124)	(0.0055)	(0.0127)
Sexual assault	-0.0085	-0.0205	0.0143**	0.0335**
	(0.0067)	(0.0161)	(0.0070)	(0.0163)
Kidnapping	-0.0063	-0.0151	0.0070	0.0164
	(0.0076)	(0.0182)	(0.0075)	(0.0176)
Murder	-0.0149	-0.0358	0.0118	0.0275
	(0.0108)	(0.0259)	(0.0117)	(0.0273)
F-stat joint F-test	3.759	3.759	2.652	2.652
P-value joint F-test	0.000	0.000	0.001	0.001
Observations	183,381	183,381	183,381	183,381

Note: This table shows balance tests where judge stringencies have been standardized to ease interpretation of magnitudes. Columns 1 and 3 show the noncarceral conviction and incarceration stringencies, standardized to have a mean of zero and a standard deviation of one. Columns 2 and 4 are similar, but the stringencies have been residualized against court-by-time fixed effects before being standardized. We regress each standardized stringency on case characteristics, controlling for court-by-year fixed effects, court-by-month 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. Star denote * p< 0.10, ** p< 0.05, *** p< 0.01.

Table D.2: Noncarceral conviction and recidivism—robustness to dropping unbalanced offenses

	Year 1		Years 2-4		Years 5-7		Years 1-7	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: No assault offenses								
Fut. charge	0.004*	0.099*	0.014***	0.067	0.011***	0.047	0.026***	0.186*
	(0.002)	(0.053)	(0.003)	(0.081)	(0.003)	(0.082)	(0.004)	(0.105)
Fut. conviction	0.006***	0.122**	0.016***	0.111	0.012***	0.040	0.028***	0.260**
	(0.002)	(0.050)	(0.003)	(0.079)	(0.003)	(0.079)	(0.004)	(0.106)
Fut. incarceration	0.005***	0.107**	0.014***	0.023	0.009***	-0.034	0.025***	0.176*
	(0.002)	(0.043)	(0.003)	(0.067)	(0.002)	(0.064)	(0.003)	(0.090)
Observations	154,138	154,138	154,138	154,138	154,138	154,138	154,138	154,138
Panel B: No drug offenses								
Fut. charge	-0.013***	0.149**	-0.010***	0.192	-0.003	0.085	-0.015***	0.339**
	(0.003)	(0.073)	(0.004)	(0.126)	(0.003)	(0.109)	(0.005)	(0.151)
Fut. conviction	-0.010***	0.209***	-0.006	0.241*	-0.001	0.086	-0.011**	0.436***
	(0.003)	(0.069)	(0.003)	(0.125)	(0.003)	(0.107)	(0.004)	(0.157)
Fut. incarceration	-0.007***	0.166***	-0.006**	0.179	-0.002	-0.044	-0.011***	0.329**
	(0.002)	(0.063)	(0.003)	(0.111)	(0.003)	(0.087)	(0.004)	(0.136)
Observations	125,541	125,541	125,541	125,541	125,541	125,541	125,541	125,541
Panel C: No violent offenses								
Fut. charge	0.004*	0.093	0.016***	0.068	0.012***	0.047	0.028***	0.190*
	(0.002)	(0.057)	(0.003)	(0.086)	(0.003)	(0.086)	(0.004)	(0.111)
Fut. conviction	0.006***	0.125**	0.018***	0.102	0.014***	0.033	0.031***	0.260**
	(0.002)	(0.054)	(0.003)	(0.084)	(0.003)	(0.082)	(0.004)	(0.112)
Fut. incarceration	0.005***	0.110**	0.016***	0.024	0.010***	-0.045	0.028***	0.177*
	(0.002)	(0.047)	(0.003)	(0.071)	(0.002)	(0.068)	(0.003)	(0.097)
Observations	149,069	149,069	149,069	149,069	149,069	149,069	149,069	149,069
Panel D: No assault, sexual assault, fraud, or traffic offenses								
Fut. charge	0.005*	0.109	0.018***	0.018	0.009***	-0.023	0.028***	0.160
	(0.003)	(0.073)	(0.004)	(0.103)	(0.003)	(0.101)	(0.004)	(0.136)
Fut. conviction	0.006***	0.160**	0.020***	0.041	0.012***	-0.067	0.032***	0.232*
	(0.002)	(0.070)	(0.003)	(0.099)	(0.003)	(0.097)	(0.004)	(0.133)
Fut. incarceration	0.006***	0.146**	0.017***	0.005	0.009***	-0.108	0.030***	0.163
	(0.002)	(0.059)	(0.003)	(0.089)	(0.002)	(0.082)	(0.004)	(0.118)
Observations	112,088	112,088	112,088	112,088	112,088	112,088	112,088	112,088

Note: This table reports 2SLS results for the impact of noncarceral conviction vs dismissal after excluding various offense categories. For Panel A, we drop assault cases, recalculate the instrument and run the 2SLS regression on all non-assault cases. Panel B is similar except it only omits drug cases. Panel C is similar except it omits all violent offenses (assault, sexual assault, and murder) and Panel D omits all offense types with evidence of imbalance (assault, sexual assault, fraud, and traffic). The columns report results for four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). All regressions control for z_i, race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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.3: Incarceration and recidivism—robustness to dropping unbalanced offenses

	Year 1		Years 2-4		Years 5-7		Years 1-7	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A: No assault offenses								
Fut. charge	-0.021***	-0.090***	0.018***	-0.001	0.028***	-0.003	0.030***	-0.052
	(0.002)	(0.031)	(0.003)	(0.047)	(0.002)	(0.042)	(0.003)	(0.059)
Fut. conviction	-0.017***	-0.102***	0.019***	-0.027	0.026***	0.013	0.029***	-0.094
	(0.002)	(0.030)	(0.002)	(0.048)	(0.002)	(0.040)	(0.003)	(0.058)
Fut. incarceration	-0.009***	-0.066***	0.021***	0.007	0.024***	0.037	0.034***	-0.048
	(0.001)	(0.025)	(0.002)	(0.042)	(0.002)	(0.033)	(0.003)	(0.052)
Observations	154,138	154,138	154,138	154,138	154,138	154,138	154,138	154,138
Panel B: No drug offenses								
Fut. charge	-0.017***	-0.117***	0.015***	0.003	0.026***	0.037	0.028***	-0.036
	(0.002)	(0.040)	(0.003)	(0.064)	(0.002)	(0.053)	(0.003)	(0.080)
Fut. conviction	-0.014***	-0.138***	0.016***	-0.030	0.024***	0.037	0.028***	-0.089
	(0.002)	(0.038)	(0.003)	(0.062)	(0.002)	(0.051)	(0.003)	(0.077)
Fut. incarceration	-0.007***	-0.084***	0.019***	0.003	0.022***	0.087**	0.034***	-0.012
	(0.002)	(0.032)	(0.002)	(0.055)	(0.002)	(0.043)	(0.003)	(0.069)
Observations	125,541	125,541	125,541	125,541	125,541	125,541	125,541	125,541
Panel C: No violent offenses								
Fut. charge	-0.020***	-0.085***	0.020***	0.013	0.029***	0.006	0.033***	-0.028
	(0.002)	(0.032)	(0.003)	(0.049)	(0.002)	(0.043)	(0.003)	(0.060)
Fut. conviction	-0.017***	-0.097***	0.020***	-0.005	0.027***	0.023	0.031***	-0.069
	(0.002)	(0.031)	(0.003)	(0.049)	(0.002)	(0.041)	(0.003)	(0.058)
Fut. incarceration	-0.009***	-0.065**	0.022***	0.025	0.025***	0.045	0.035***	-0.029
	(0.001)	(0.026)	(0.002)	(0.043)	(0.002)	(0.034)	(0.003)	(0.053)
Observations	149,069	149,069	149,069	149,069	149,069	149,069	149,069	149,069
Panel D: No assault, sexual assault, fraud, or traffic offenses								
Fut. charge	-0.021***	-0.077*	0.022***	0.033	0.033***	0.038	0.036***	-0.005
	(0.002)	(0.041)	(0.003)	(0.059)	(0.003)	(0.055)	(0.003)	(0.076)
Fut. conviction	-0.017***	-0.097**	0.022***	0.024	0.030***	0.082	0.035***	-0.030
	(0.002)	(0.040)	(0.003)	(0.058)	(0.002)	(0.052)	(0.003)	(0.073)
Fut. incarceration	-0.009***	-0.065*	0.023***	0.040	0.026***	0.054	0.036***	-0.005
	(0.002)	(0.035)	(0.003)	(0.055)	(0.002)	(0.043)	(0.003)	(0.067)
Observations	112,088	112,088	112,088	112,088	112,088	112,088	112,088	112,088

Note: This table reports 2SLS results for the impact of incarceration vs noncarceral conviction after excluding various offense categories. For Panel A, we drop assault cases, recalculate the instrument and run the 2SLS regression on all non-assault cases. Panel B is similar except it only omits drug cases. Panel C is similar except it omits all violent offenses (assault, sexual assault, and murder) and Panel D omits all offense types with evidence of imbalance (assault, sexual assault, fraud, and traffic). The columns report results for four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). All regressions control for z_d, race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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.4: 2SLS regressions of sentence length on conviction stringency

	Sent. length	Any incar.	6mo	1y	2y	3y	4y	5y	6y	7y
Pr. convict	8.48	-0.032	0.090*	-0.031	-0.043	-0.027	0.0024	-0.011	-0.0071	-0.00013
	(64.3)	(0.051)	(0.047)	(0.043)	(0.033)	(0.029)	(0.025)	(0.021)	(0.017)	(0.017)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	322.018	0.546	0.374	0.203	0.113	0.078	0.061	0.042	0.035	0.030
N	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table shows a regression of various sentence length variables on z_c, controlling for z_i. The first column uses sentence length as the outcome, the second column uses any incarceration, and columns 3–10 use any incarceration greater than 6 months, 1 year, 2 years, 3 years, 4 years, 5 years, 6 years, or 7 years, respectively. All regressions control for z_i, race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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.5: Split sample monotonicity test

	Conviction		Incarceration	
	Zero	One	Zero	One
Any drug charges	0.546	0.199	0.537	0.366
Any property charges	0.459	0.238	0.676	0.342
Any violent charges	0.430	0.098	0.318	0.191
Black	0.306	0.392	0.460	0.592
Female	0.875	0.168	0.658	0.269
Prior conviction	0.269	0.149	0.737	0.350

Note: This table shows first-stage estimates for the conviction and incarceration instruments 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 cases without drug charges, then estimates the first stage on cases *with* drug charges, and reports the coefficient on the instrument. Regression includes court-by-year fixed effects, court-by-month fixed effects, and day-of-week fixed effects. Standard errors are clustered at the judge-year level. This 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

E.1.1 Disposition types

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 for each offense type.

Future exposure to incarceration. Appendix Figure E.2 illustrates the extent of "incarceration catch-up" for individuals given noncarceral sentences compared to those given carceral sentences, considering both new crimes and technical violations leading to probation revocation. Although there is some catch-up, over 50% of individuals receiving noncarceral sentences avoid incarceration over the next seven years.

E.1.2 Reduced-form estimates

Panel A of 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 fixed effects, and day-of-week fixed effects as well as the leave-one-out judge incarceration stringency. Our conviction instrument positively and significantly affects the year 1 and the years 1–7 outcomes. Panel B of Appendix Table E.1 shows comparable reduced-form estimates for the incarceration results.

E.1.3 Compliers

Characterizing compliers. Appendix Table E.2 compares compliers for the conviction and incarceration margins to the full sample. Compliers to the conviction instrument are more likely to be female (27% vs 22%) relative to the overall sample and are more likely to have a property crime charge (42% vs 38%). They are less likely to have a prior conviction (10% vs 17%), less likely to have a violent charge (8% vs 19%), and less likely to have charges that fall into the "other" category (6% vs 16%). Compliers to the incarceration instrument are more similar to the full sample. For instance, prior conviction rates and share of women are near identical: 18% vs 17% and 21% vs 22%, respectively. For property charges and violent charges we see some disparities, with 46% of complier cases having a property charge vs 38% of overall cases, and 8% of compliers having a violent charge vs 19% overall.

Complier weighted OLS. In Appendix Table E.3 we reweight the OLS for incarceration and conviction margin compliers. The OLS estimates do not change much when re-weighting for compliers. The reweighted estimates for noncarceral conviction are somewhat larger, while the estimates for incarceration are nearly identical.

E.1.4 **Heterogeneity**

Increased criminal behavior or "ratcheting up"?   We take two strategies to provide suggestive evidence on whether the recidivism effects come from increased criminal behavior or the "ratcheting up" effect. First, we consider outcomes in different stages of the criminal justice process. If each discretionary decision is influenced by the criminal record, then the influence of the conviction 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 a ratcheting up mechanism, in all of our estimates presented in Table IV the percent changes are larger for more downstream measures of future criminal justice contact.[11]

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.[12] Seeing that our results are driven by an increase in income-generating crime would provide support for the destabilization channel. Appendix Table E.4 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. Our results are similar if we break out drug crimes from non-drug crimes (Appendix Table E.5).

2SLS estimates for other subgroups.   In Appendix Tables E.6–E.8, we present 2SLS estimates conditional on various offense categories and sociodemographic characteristics. Appendix Table E.6 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 the standard errors are too large for us to determine differential impacts across groups.

For incarceration, we find that both groups have similar patterns: short-term incapacitation effects, but no long-term effects for either group. This result differs from findings in Jordan, Karger, and Neal (2023). This difference could partially be caused by two limitations in our data. First, we can only observe prior felony convictions if they appear in our data set. Our indicator for prior felony conviction is "prior felony conviction within the last 5 years" (and would miss all felony convictions outside of the state). Presumably, some subset of our sample with no felony conviction within the last five years have older felony convictions we cannot observe. Jordan, Karger, and Neal (2023) solve this issue by restricting their analysis to individuals who are younger than 18 at the start of their sample. We are not able to include a similar restriction because we do not know the age or date of birth for many people in our sample. It is possible that we would find different results for incarceration if our data allowed us to fully restrict the sample to first-time offenders.

We find no substantial differences between Black and White defendants (Appendix Table E.7). We do find some evidence that impacts are larger for people living in zip codes with

---

[11]The fact that conviction increases the probability of future incarceration also indicates that there these marginal convictions impose direct future costs on the criminal justice system.

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

above median poverty rates (Appendix Table E.8). One interpretation could be that felony convictions have greater consequences in terms of access to social services or housing, or in terms of future criminal justice scrutiny, for poorer people.

E.1.5 **Robustness checks**

Robustness to sample choice and specification. In Appendix Figures E.3–E.6, we examine how our main 2SLS estimates for conviction and incarceration change when we alter our sample or specifications, for our 1 year, 2–4 years, 5–7 years, and 1–7 years estimates. We consider the following variations:

- Changing the required number of cases seen by a judge in our three-year window (50 or 150 instead of 100).

- Varying which courts are included. We conducted phone interviews in 2021 with court clerks in all courts in Virginia for which we had data. We asked the clerks how cases were allocated. Our main sample includes courts where cases are quasi-randomly allocated (see Section *II.A* for more details.) We vary which courts we include:

  - Keep all courts, even if there appears to be selection in the kinds of cases that judges handle, e.g. all violent cases go to one judge and nonviolent cases go to another.
  - In addition to the sample restrictions from the main analyses, drop courts where the clerks said that cases were assigned based on judge availability, which may be more subject to discretion in what cases to work on.

- Clustering our standard errors at the month court level or at the defendant level.

- Changing what offenses are included:

  - Dropping drug cases. Although diversion is rare for felonies in Virginia, it is more likely in drug cases. Dropping drug cases means eliminating the cases where diversion is most probable.
  - Dropping offense types that are not balanced across judges (see Table III).

- Varying how we control for non-focal stringency. In our main specification, we control for incarceration stringency, defined as the fraction of cases that end in carceral sentences. Here, we consider several regressions including (in addition to our baseline controls), controls for (1) sentence length stringency, (2) probability of sentence length shorter than 6 months,(3) sentence length longer than 1 year or (4) longer than 4 years,and (5) flexibly controlling for deciles of the non-focal stringency.

- Reconstructing the judge stringency instrument by crime type (assault or not and drug or not).

- Including all years for which we can construct recidivism. We expand the sample up to 2015 for outcomes in years 2–4 and up to 2018 for outcomes in year 1.

- Removing all case and individual controls from the regression.

Generally, our estimates are very close to the results from 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.

Robustness to different definitions of recidivism. In Appendix Table E.9, we show that our results are robust to how we define recidivism. In panel A we count recidivism as the total number of future charges (i.e., if you have 3 future charges in a case 1 year later, we count that as 3.) In panel B we count the total number of future charge events (i.e., if you have a case in year 1 and another separate case in year 2 we count that as 2). Finally in panels C, D, and E we look at recidivism where there is one charge, two to three charges, or four or more charges. While we see the same general patterns, our estimates occasionally fall into or out of significance. Furthermore, much of our results seem to be coming from recidivism with more than one charge, as evidenced from panels D and E.

Empirical Bayes Shrinkage. We correct for potential measurement error in judge stringency instruments using Empirical Bayes methods. We implement an Empirical Bayes procedure where we assume that judge stringencies are drawn from a Beta distribution, and the individual stringencies follow a Bernoulli distribution. We consider two specifications: in the first, we assume that judge stringencies are drawn from a single Beta distribution, while the second assumes that the Beta distribution varies by circuit-year. We provide detailed descriptions of our methodology and results in Appendix E.4. Overall, our results are not sensitive to using shrunken leniency estimates, which is consistent with the fact that judges in our sample see many cases per year.

Differential mobility. Our results could be confounded if conviction or incarceration influence the likelihood of moving outside of Virginia, and therefore change 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 (described in more detail in Appendix Subsection G), 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 or year 7 post-sentencing. We exclude missing observations.[13] As we can see in Appendix Figure E.7, there is no discontinuity at our cutoff score. 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. This consistency suggests that neither conviction nor incarceration affect migration from Virginia.

---

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

E.2 Appendix figures: 2SLS analyses

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

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

Figure E.2: Dynamics of incarceration

(a) Ever incarcerated with or without technical revocations

(b) Ever incarcerated with or without technical revocations—complier reweighted
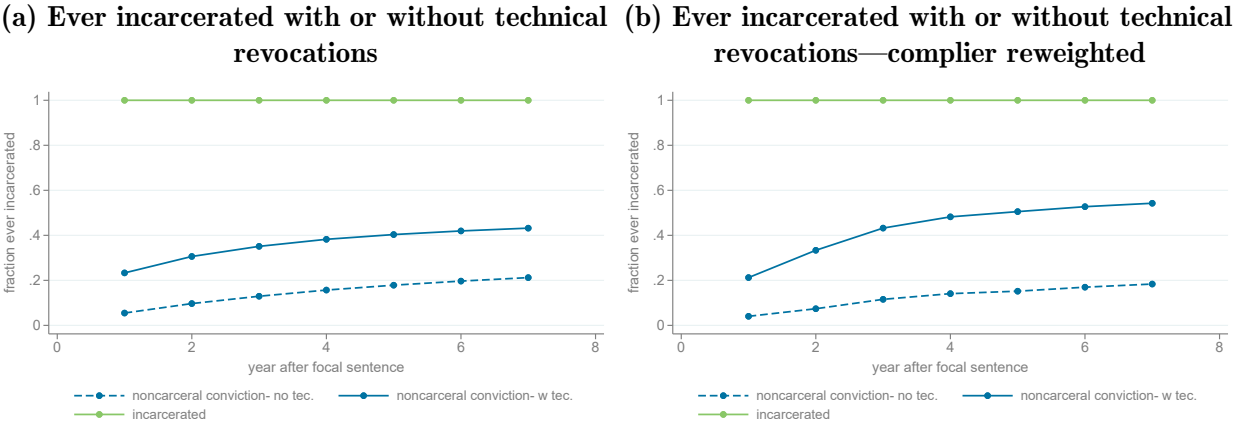
Note: Panel (a) shows the fraction of individuals that currently or previously have been incarcerated in each year after the focal sentence date. This takes into account the current sentence, prior incarceration (as recorded in the sentence guidelines worksheet), and any future carceral sentences. The green line shows those incarcerated in the focal sentence. The blue dotted line shows incarceration due to new criminal activity for those who received a noncarceral conviction, and the solid blue line shows incarceration both for new criminal activity and for technical violations for those who received a noncarceral conviction. Panel (b) is similar but complier-weighted, meaning that it is reweighted to match the compliers for each margin.

Figure E.3: Robustness for 2SLS results: recidivism in year 1
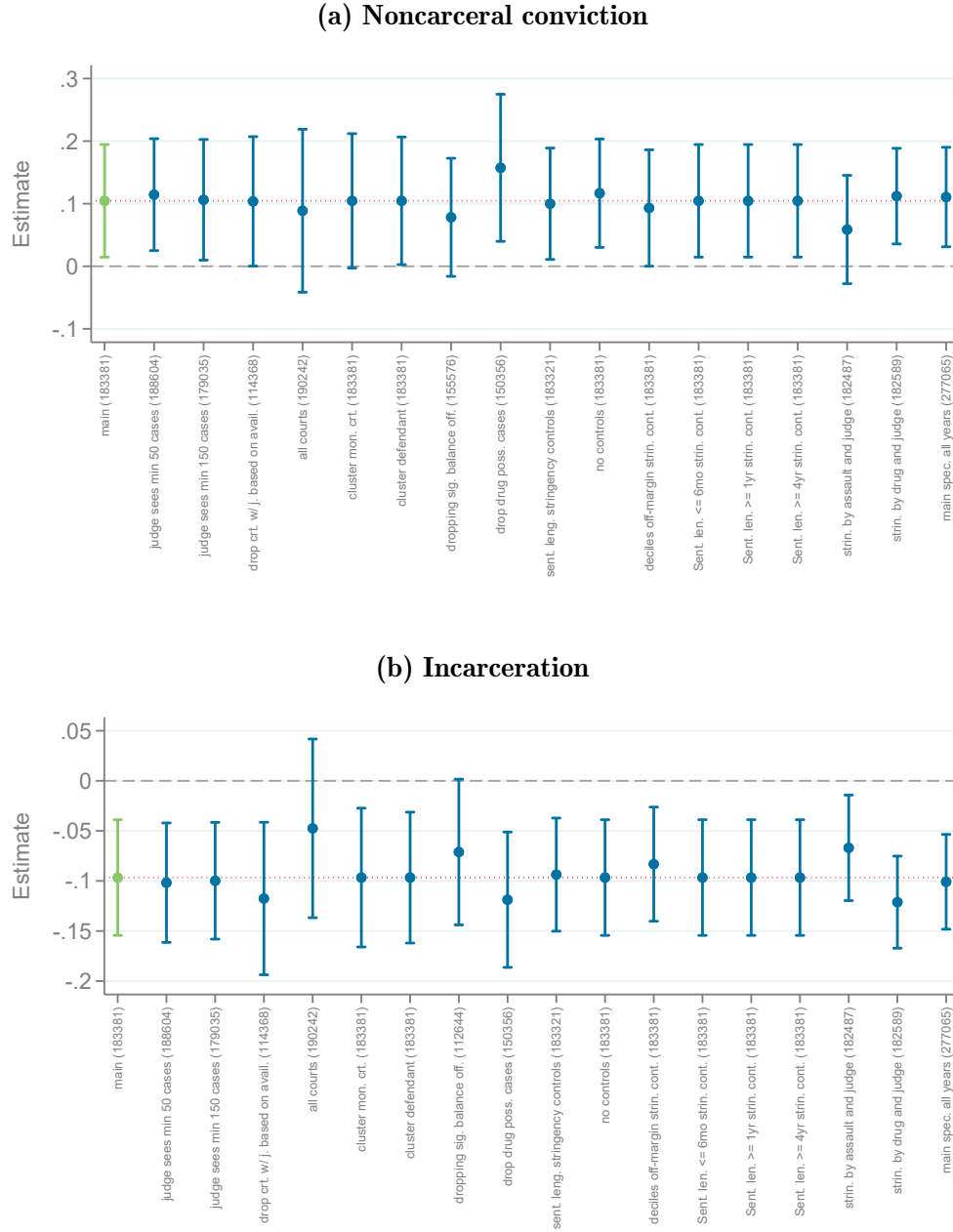
(a) Noncarceral conviction

(b) Incarceration

Note: This figure shows various estimates of the impact of conviction (panel a) and incarceration (panel b) on the likelihood of a new felony charge 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. We consider several different specifications: (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 assign judges based on availability. (5) Keeping courts where clerks described a non-random assignment process. (6) Clustering standard errors at the court-month level. (7) Clustering standard errors at the defendant level. (8) Dropping any offenses that are significant in our balance tests. (9) Dropping any cases that relate to drug possession. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls. (12) Including decile bins of our off-margin stringency as controls. (13) Controlling for "probability of sentence length less than 6 months" stringency. (14) Controlling for "probability of sentence length greater than 1 year" stringency. (15) Controlling for "probability of sentence length greater than 4 years" stringency. (16) Controlling for judge stringency instruments recalculated by assault only/non-assault charges. (17) Controlling for judge stringency instruments recalculated by drug/non-drug cases. (18) Using the full sample of available years for the estimate.

Figure E.4: Robustness for 2SLS results: recidivism in years 2–4
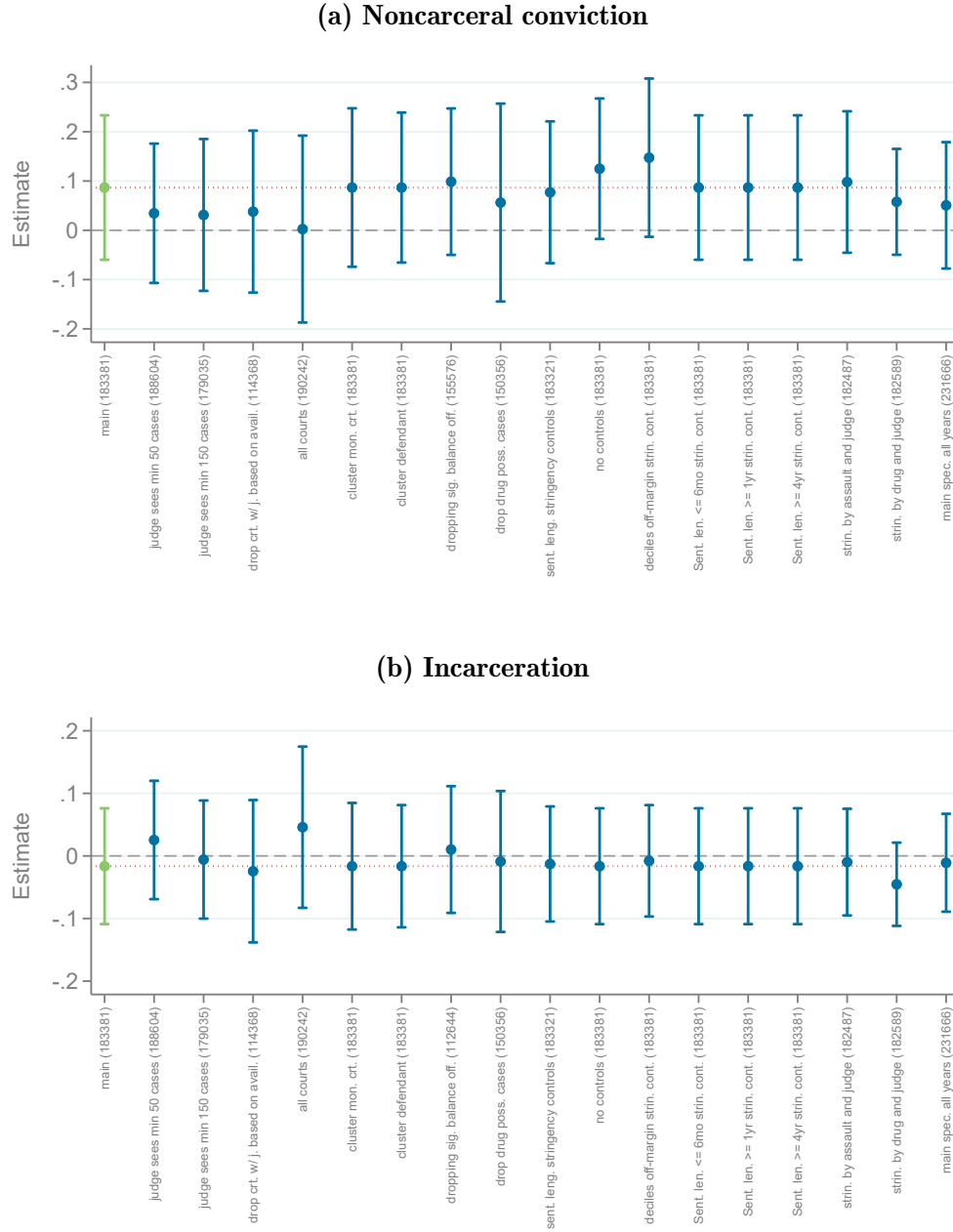
(a) Noncarceral conviction

(b) Incarceration

Note: This figure shows various estimates of the impact of conviction (panel a) and incarceration (panel b) on the likelihood of a new felony charge 2–4 years 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. We consider several different specifications: (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 a non-random assignment process. (6) Clustering standard errors at the court-month level. (7) Clustering standard errors at the defendant level. (8) Dropping any offenses that are significant in our balance tests. (9) Dropping any cases that relate to drug possession. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls. (12) Including decile bins of our off-margin stringency as controls. (13) Controlling for "probability of sentence length less than 6 months" stringency. (14) Controlling for "probability of sentence length greater than 1 year" stringency. (15) Controlling for "probability of sentence length greater than 4 years" stringency. (16) Controlling for judge stringency instruments recalculated by assault only/non-assault charges. (17) Controlling for judge stringency instruments recalculated by drug/non-drug cases. (18) Using the full sample of available years for the estimate.

Figure E.5: Robustness for 2SLS results: recidivism in years 5–7
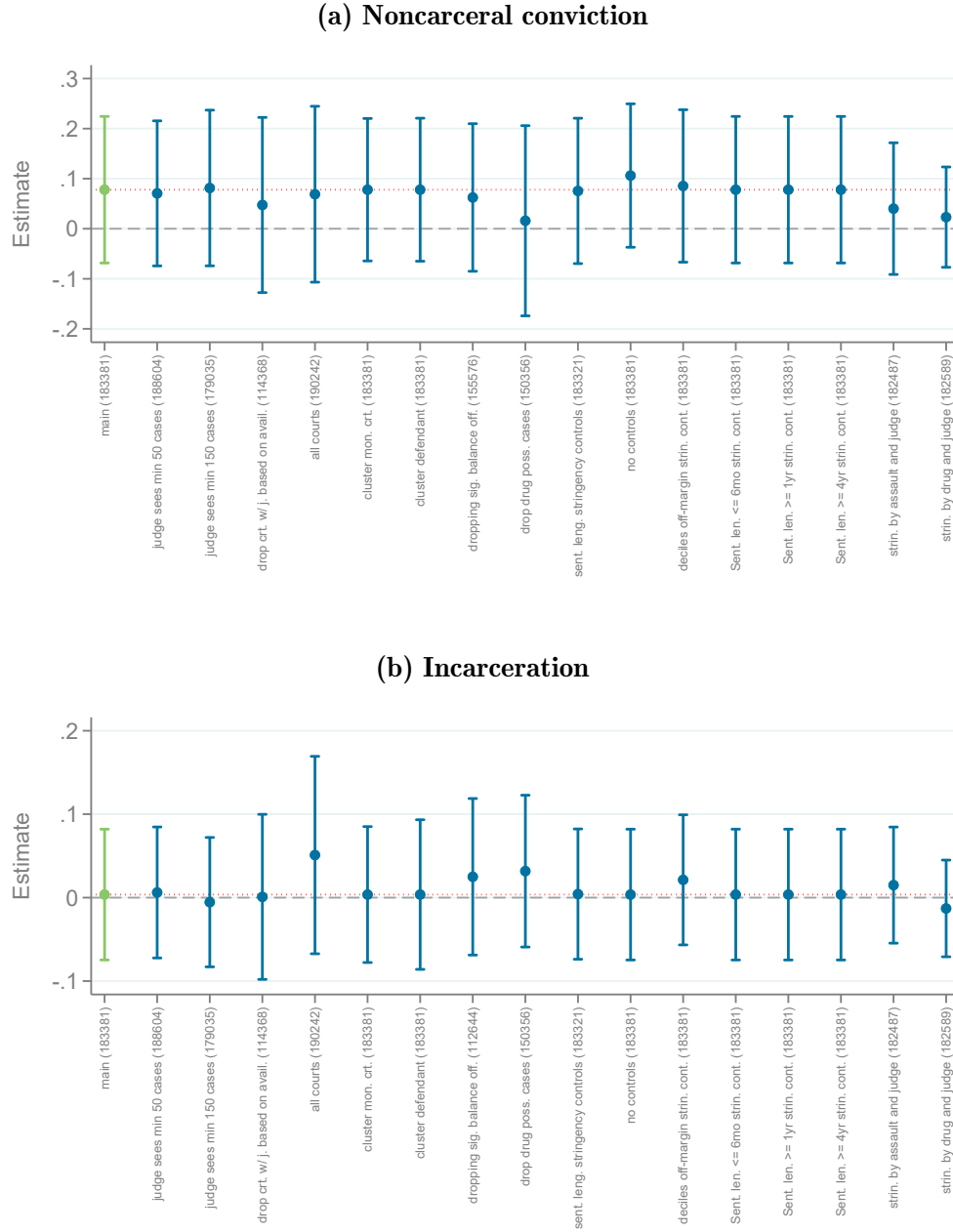
(a) Noncarceral conviction

(b) Incarceration

Note: This figure shows various estimates of the impact of conviction (panel a) and incarceration (panel b) on the likelihood of a new felony charge 5–7 years 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. We consider several different specifications: (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 a non-random assignment process. (6) Clustering standard errors at the court-month level. (7) Clustering standard errors at the defendant level. (8) Dropping any offenses that are significant in our balance tests. (9) Dropping any cases that relate to drug possession. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls. (12) Including decile bins of our off-margin stringency as controls. (13) Controlling for "probability of sentence length less than 6 months" stringency. (14) Controlling for "probability of sentence length greater than 1 year" stringency. (15) Controlling for "probability of sentence length greater than 4 years" stringency. (16) Controlling for judge stringency instruments recalculated by assault only/non-assault charges. (17) Controlling for judge stringency instruments recalculated by drug/non-drug cases.

Figure E.6: Robustness for 2SLS results: recidivism in years 1-7

(a) Noncarceral conviction

(b) Incarceration

Note: This figure shows various estimates of the impact of conviction (panel a) and incarceration (panel b) on the likelihood of a new felony charge within the first seven years 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. We consider several different specifications: (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 aa non-random assignment process. (6) Clustering standard errors at the court-month level. (7) Clustering standard errors at the defendant level. (8) Dropping any offenses that are significant in our balance tests. (9) Dropping any cases that relate to drug possession. (10) Including a sentence length stringency instrument control. (11) Main specification without any of our controls. (12) Including decile bins of our off-margin stringency as controls. (13) Controlling for "probability of sentence length less than 6 months" stringency. (14) Controlling for "probability of sentence length greater than 1 year" stringency. (15) Controlling for "probability of sentence length greater than 4 years" stringency. (16) Controlling for judge stringency instruments recalculated by assault only/non-assault charges. (17) Controlling for judge stringency instruments recalculated by drug/non-drug cases.

Figure E.7: Testing for discontinuities in Virginia residency

(a) Incarceration-length sample　　　　**(b) Probation/jail sample**
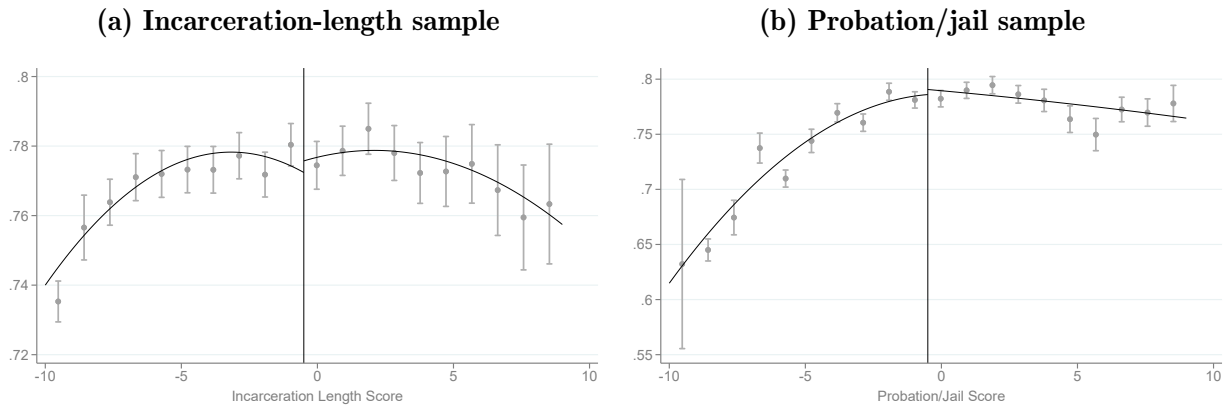
Note: These figures check for discontinuities in the likelihood of being a Virginia resident as a function of the sentence guidelines score. Those who score above the cutoff of 0 in panel (a) receive a sentence that is approximately 8 months longer than those who score right below the cutoff. Those who score above the cutoff of 0 in panel (b) are about 40 percentage points more likely to receive a short jail sentence than those right below, who mostly receive probation. The samples for the regression discontinuity designs are described in Appendix Section G. The outcome variable 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. We excluded individuals whose residency information is missing (7.7% of the sample) from the analysis.

E.3 Appendix tables: 2SLS analyses

Table E.1: Reduced form estimates

| | Year 1 | Years 2-4 | Years 5-4 | Years 1-7 |
	RF	RF	RF	RF
Panel A: Conviction				
Fut. charge	0.062**	0.051	0.046	0.139**
	(0.027)	(0.044)	(0.043)	(0.055)
Fut. conviction	0.080***	0.068	0.033	0.177***
	(0.025)	(0.042)	(0.042)	(0.053)
Fut. incarceration	0.067***	0.035	-0.014	0.127***
	(0.022)	(0.037)	(0.034)	(0.048)
Observations	183,381	183,381	183,381	183,381
Panel B: Incarceration				
Fut. charge	-0.058***	-0.010	0.002	-0.043
	(0.018)	(0.028)	(0.024)	(0.036)
Fut. conviction	-0.067***	-0.023	0.013	-0.064*
	(0.017)	(0.028)	(0.023)	(0.035)
Fut. incarceration	-0.043***	0.004	0.032	-0.018
	(0.014)	(0.025)	(0.019)	(0.031)
Observations	183,381	183,381	183,381	183,381

Note: This table shows estimates from reduced form regressions of recidivism on z_c in Panel A and regressions of recidivism on z_i in Panel B. The four columns report results for four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). All regressions in the first panel control for z_i and all in the second panel control for z_d. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month fixed effects, and day-of-week fixed effects. 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: Complier characteristics (noncarceral conviction)

	$Pr(X=x)$	$Pr(X=x\|complier)$	$\frac{Pr(X=x\|complier)}{Pr(X=x)}$
Panel A: Conviction			
Prior conviction	0.173	0.101	0.584
	(0.003)	(0.031)	(0.180)
Female	0.218	0.273	1.248
	(0.003)	(0.041)	(0.186)
Black	0.568	0.557	0.980
	(0.015)	(0.049)	(0.085)
Has misdemeanor	0.078	0.080	1.024
	(0.004)	(0.020)	(0.254)
Drugs	0.313	0.316	1.011
	(0.007)	(0.034)	(0.105)
Property	0.377	0.417	1.104
	(0.008)	(0.045)	(0.115)
Violent	0.194	0.084	0.433
	(0.004)	(0.031)	(0.158)
Other	0.160	0.064	0.397
	(0.002)	(0.027)	(0.170)
Panel B: Incarceration			
Prior conviction	0.173	0.183	1.059
	(0.003)	(0.021)	(0.117)
Female	0.218	0.209	0.956
	(0.003)	(0.031)	(0.140)
Black	0.568	0.549	0.966
	(0.015)	(0.029)	(0.046)
Has misdemeanor	0.078	0.061	0.787
	(0.004)	(0.018)	(0.222)
Drugs	0.313	0.264	0.845
	(0.007)	(0.028)	(0.088)
Property	0.377	0.460	1.220
	(0.008)	(0.034)	(0.091)
Violent	0.194	0.084	0.431
	(0.004)	(0.028)	(0.139)
Other	0.160	0.150	0.937
	(0.002)	(0.023)	(0.144)

Note: This table shows the characteristics of compliers for our 2SLS conviction analysis in Panel A and incarceration analysis in Panel B. The first column reports average characteristics for the full 2SLS sample. The second column reports the estimated average characteristics 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.3: Complier weighted OLS

	Year 1		Years 2-4		Years 5-7		Years 1-7	
	OLS	OLS weighted	OLS	OLS weighted	OLS	OLS weighted	OLS	OLS weighted
Panel A: Conviction								
Fut. charge	-0.002	0.004**	0.004	0.010***	0.006**	0.010***	0.011***	0.021***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)
Fut. conviction	0.001	0.006***	0.008***	0.012***	0.007***	0.011***	0.014***	0.023***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
Fut. incarceration	0.001	0.006***	0.006**	0.010***	0.005**	0.008***	0.012***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Ctrl. mean: fut. chrg.	0.081	0.089	0.154	0.170	0.115	0.129	0.270	0.297
Ctrl. mean: fut. conv.	0.068	0.076	0.134	0.148	0.101	0.114	0.242	0.268
Ctrl. mean: fut. incar.	0.047	0.054	0.097	0.109	0.073	0.083	0.181	0.204
Panel B: Incarceration								
Fut. charge	-0.022***	-0.022***	0.013***	0.013***	0.025***	0.025***	0.022***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Fut. conviction	-0.018***	-0.018***	0.013***	0.014***	0.023***	0.023***	0.022***	0.023***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Fut. incarceration	-0.010***	-0.010***	0.017***	0.017***	0.021***	0.021***	0.027***	0.028***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Ctrl. mean: fut. chrg.	0.088	0.088	0.177	0.175	0.133	0.132	0.308	0.306
Ctrl. mean: fut. conv.	0.078	0.077	0.160	0.159	0.121	0.120	0.285	0.283
Ctrl. mean: fut. incar.	0.055	0.055	0.116	0.115	0.085	0.084	0.214	0.212
Observations	183,381		183,381		183,381		183,381	

Note: This table shows both regular OLS and complier-weighted OLS estimates of the impact of conviction and incarceration on recidivism. The first three rows of each panel measure recidivism as any future charge, any future conviction, and any future incarceration. Recidivism is measured from the time of sentencing and within the time windows shown at the top of the table. For the regular OLS estimates we regress our measures of recidivism on dummies for conviction (regardless of incarceration status) and incarceration. For the complier-weighted OLS estimates, we do the same but each observation is weighted by the likelihood of being an IV complier. These weights differ between Panel A and Panel B as they consider different compliers. All regressions control for race, gender, prior conviction, offense type dummies, year-by-court fixed effects, court-by-month 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: Income-generating vs non-income-generating recidivism

	Income generating recidivsim				Non-income generating recidivsim			
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Conviction								
Fut. charge	0.061*	0.032	0.042	0.131*	0.091**	0.036	0.007	0.129
	(0.034)	(0.061)	(0.054)	(0.077)	(0.039)	(0.063)	(0.062)	(0.082)
Fut. conviction	0.070**	0.096	0.042	0.197**	0.100***	0.033	-0.019	0.119
	(0.032)	(0.060)	(0.052)	(0.077)	(0.036)	(0.058)	(0.060)	(0.081)
Fut. incarceration	0.046*	0.046	-0.016	0.097	0.091***	0.041	-0.030	0.133**
	(0.027)	(0.050)	(0.045)	(0.068)	(0.030)	(0.049)	(0.045)	(0.066)
Ctrl. comp. mean: fut. chrg.	0.100	0.229	0.163	0.386	0.079	0.152	0.131	0.293
Ctrl. mean: fut. chrg.	0.054	0.106	0.079	0.196	0.053	0.108	0.081	0.204
Ctrl. comp. mean: fut. conv.	0.089	0.203	0.153	0.359	0.065	0.122	0.119	0.256
Ctrl. mean: fut. conv.	0.047	0.092	0.070	0.175	0.044	0.091	0.070	0.178
Ctrl. comp. mean: fut. incar.	0.090	0.211	0.189	0.401	0.064	0.141	0.139	0.288
Ctrl. mean: fut. incar.	0.034	0.069	0.053	0.135	0.030	0.064	0.048	0.128
Panel B: Incarceration								
Fut. charge	-0.043*	0.003	-0.020	-0.054	-0.064***	-0.023	0.028	-0.032
	(0.024)	(0.036)	(0.031)	(0.048)	(0.024)	(0.041)	(0.032)	(0.051)
Fut. conviction	-0.053**	-0.014	-0.027	-0.073	-0.071***	-0.019	0.054*	-0.017
	(0.023)	(0.036)	(0.030)	(0.047)	(0.023)	(0.039)	(0.031)	(0.049)
Fut. incarceration	-0.032*	-0.003	0.025	-0.018	-0.043**	0.009	0.059**	0.024
	(0.019)	(0.031)	(0.026)	(0.042)	(0.019)	(0.032)	(0.024)	(0.040)
Ctrl. comp. mean: fut. chrg.	0.090	0.125	0.108	0.259	0.066	0.131	0.084	0.236
Ctrl. mean: fut. chrg.	0.056	0.112	0.080	0.204	0.047	0.102	0.080	0.195
Ctrl. comp. mean: fut. conv.	0.063	0.104	0.092	0.215	0.045	0.109	0.060	0.190
Ctrl. mean: fut. conv.	0.049	0.102	0.073	0.188	0.041	0.090	0.071	0.174
Ctrl. comp. mean: fut. incar.	0.036	0.049	0.040	0.119	0.020	0.040	0.028	0.088
Ctrl. mean: fut. incar.	0.035	0.074	0.053	0.141	0.029	0.062	0.047	0.123
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: The first panel of this table shows 2SLS estimates of the impact of conviction vs dismissal on future recidivism. In the first four columns, recidivism is defined in reference to new income-generating felony charges; in the last four 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 four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). 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 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: Drug vs non-drug recidivism

	Drug charges				Non-drug charges			
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Conviction								
Fut. charge	0.143**	0.006	-0.013	0.164	0.078	0.148	0.141	0.299**
	(0.070)	(0.106)	(0.110)	(0.141)	(0.057)	(0.097)	(0.086)	(0.120)
Fut. conviction	0.112*	0.031	-0.050	0.207	0.144***	0.173*	0.124	0.370***
	(0.062)	(0.099)	(0.104)	(0.134)	(0.054)	(0.093)	(0.083)	(0.118)
Fut. incarceration	0.115**	-0.026	-0.082	0.133	0.111**	0.110	0.019	0.273***
	(0.056)	(0.091)	(0.087)	(0.129)	(0.048)	(0.083)	(0.068)	(0.103)
Ctrl. comp. mean: fut. chrg.	0.148	0.356	0.250	0.554	0.164	0.272	0.229	0.459
Ctrl. mean: fut. chrg.	0.079	0.159	0.123	0.282	0.094	0.176	0.132	0.306
Ctrl. comp. mean: fut. conv.	0.133	0.326	0.233	0.521	0.141	0.231	0.221	0.424
Ctrl. mean: fut. conv.	0.067	0.137	0.108	0.252	0.080	0.154	0.117	0.277
Ctrl. comp. mean: fut. incar.	0.135	0.333	0.281	0.570	0.137	0.267	0.279	0.503
Ctrl. mean: fut. incar.	0.047	0.097	0.076	0.184	0.058	0.116	0.087	0.216
Panel B: Incarceration								
Fut. charge	-0.062	-0.007	-0.015	-0.056	-0.112***	-0.025	0.005	-0.087
	(0.061)	(0.089)	(0.082)	(0.110)	(0.034)	(0.055)	(0.046)	(0.069)
Fut. conviction	-0.075	-0.007	0.027	-0.092	-0.128***	-0.054	0.014	-0.121*
	(0.055)	(0.089)	(0.079)	(0.109)	(0.033)	(0.053)	(0.045)	(0.066)
Fut. incarceration	-0.066	0.077	0.008	-0.025	-0.076***	-0.020	0.062*	-0.038
	(0.050)	(0.080)	(0.068)	(0.103)	(0.028)	(0.046)	(0.037)	(0.058)
Ctrl. comp. mean: fut. chrg.	0.123	0.178	0.105	0.341	0.118	0.205	0.162	0.377
Ctrl. mean: fut. chrg.	0.100	0.189	0.143	0.336	0.082	0.168	0.126	0.291
Ctrl. comp. mean: fut. conv.	0.084	0.132	0.065	0.260	0.083	0.181	0.133	0.328
Ctrl. mean: fut. conv.	0.088	0.171	0.130	0.311	0.072	0.153	0.114	0.269
Ctrl. comp. mean: fut. incar.	0.043	0.048	0.026	0.136	0.043	0.083	0.064	0.180
Ctrl. mean: fut. incar.	0.063	0.125	0.090	0.234	0.051	0.110	0.081	0.201
Observations	57,249	57,249	57,249	57,249	126,134	126,134	126,134	126,134

Note: The first panel of this table shows 2SLS estimates of the impact of conviction vs dismissal on future recidivism. In the first four columns, recidivism is defined in reference to new drug charges; in the last four 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 four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). 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 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 those with/without prior felony convictions

	Priors				No priors			
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Conviction								
Fut. charge	0.041	0.411	-0.223	0.391	0.106**	0.062	0.117*	0.224**
	(0.208)	(0.364)	(0.398)	(0.438)	(0.046)	(0.073)	(0.070)	(0.093)
Fut. conviction	0.166	0.340	-0.251	0.424	0.129***	0.099	0.090	0.289***
	(0.195)	(0.343)	(0.379)	(0.443)	(0.044)	(0.071)	(0.066)	(0.090)
Fut. incarceration	0.140	0.428	-0.385	0.475	0.109***	0.025	0.010	0.188**
	(0.178)	(0.315)	(0.329)	(0.425)	(0.036)	(0.060)	(0.053)	(0.078)
Ctrl. comp. mean: fut. chrg.	0.341	0.506	0.543	0.944	0.138	0.270	0.189	0.425
Ctrl. mean: fut. chrg.	0.146	0.294	0.239	0.503	0.080	0.150	0.112	0.265
Ctrl. comp. mean: fut. conv.	0.270	0.425	0.508	0.839	0.124	0.237	0.181	0.400
Ctrl. mean: fut. conv.	0.128	0.265	0.217	0.471	0.067	0.129	0.098	0.236
Ctrl. comp. mean: fut. incar.	0.286	0.506	0.706	1.070	0.119	0.259	0.221	0.452
Ctrl. mean: fut. incar.	0.097	0.210	0.173	0.386	0.047	0.093	0.069	0.176
Panel B: Incarceration								
Fut. charge	-0.086	-0.055	0.097	-0.040	-0.092***	-0.006	-0.012	-0.068
	(0.072)	(0.112)	(0.114)	(0.132)	(0.032)	(0.049)	(0.041)	(0.061)
Fut. conviction	-0.137*	-0.113	0.101	-0.132	-0.101***	-0.020	0.007	-0.092
	(0.070)	(0.107)	(0.111)	(0.128)	(0.031)	(0.049)	(0.040)	(0.059)
Fut. incarceration	-0.082	-0.092	0.145	-0.050	-0.064**	0.031	0.038	-0.014
	(0.065)	(0.099)	(0.096)	(0.122)	(0.025)	(0.042)	(0.034)	(0.052)
Ctrl. comp. mean: fut. chrg.	0.094	0.204	0.240	0.446	0.128	0.202	0.129	0.357
Ctrl. mean: fut. chrg.	0.117	0.301	0.233	0.495	0.084	0.161	0.120	0.285
Ctrl. comp. mean: fut. conv.	0.084	0.190	0.196	0.397	0.085	0.167	0.099	0.297
Ctrl. mean: fut. conv.	0.105	0.280	0.216	0.470	0.074	0.145	0.109	0.261
Ctrl. comp. mean: fut. incar.	0.049	0.128	0.114	0.261	0.043	0.065	0.043	0.154
Ctrl. mean: fut. incar.	0.075	0.215	0.160	0.373	0.053	0.103	0.075	0.194
Observations	31,731	31,731	31,731	31,731	151,652	151,652	151,652	151,652

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 four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). 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 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 Black and non-Black defendants

	Black				Non-Black			
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Conviction								
Fut. charge	0.101	0.075	0.114	0.241	0.102	0.115	0.042	0.240**
	(0.069)	(0.117)	(0.110)	(0.147)	(0.063)	(0.090)	(0.094)	(0.115)
Fut. conviction	0.124*	0.115	0.102	0.339**	0.141**	0.128	0.011	0.272**
	(0.067)	(0.110)	(0.102)	(0.146)	(0.058)	(0.088)	(0.092)	(0.111)
Fut. incarceration	0.165***	0.010	-0.011	0.264**	0.058	0.114	-0.027	0.180*
	(0.061)	(0.088)	(0.080)	(0.126)	(0.049)	(0.082)	(0.077)	(0.103)
Ctrl. comp. mean: fut. chrg.	0.150	0.302	0.222	0.492	0.154	0.259	0.216	0.432
Ctrl. mean: fut. chrg.	0.104	0.196	0.148	0.339	0.070	0.135	0.104	0.241
Ctrl. comp. mean: fut. conv.	0.135	0.259	0.212	0.459	0.132	0.234	0.209	0.406
Ctrl. mean: fut. conv.	0.088	0.169	0.129	0.305	0.059	0.120	0.093	0.218
Ctrl. comp. mean: fut. incar.	0.136	0.323	0.290	0.566	0.128	0.232	0.249	0.445
Ctrl. mean: fut. incar.	0.064	0.126	0.094	0.235	0.040	0.087	0.069	0.164
Panel B: Incarceration								
Fut. charge	-0.131***	-0.012	-0.061	-0.116	-0.055	-0.021	0.073	-0.024
	(0.042)	(0.070)	(0.060)	(0.086)	(0.040)	(0.065)	(0.057)	(0.082)
Fut. conviction	-0.125***	-0.028	-0.045	-0.143*	-0.097**	-0.047	0.092*	-0.069
	(0.041)	(0.068)	(0.057)	(0.085)	(0.040)	(0.064)	(0.056)	(0.079)
Fut. incarceration	-0.105***	0.011	0.016	-0.072	-0.030	0.005	0.089*	0.018
	(0.035)	(0.056)	(0.047)	(0.074)	(0.032)	(0.057)	(0.049)	(0.072)
Ctrl. comp. mean: fut. chrg.	0.164	0.226	0.174	0.435	0.097	0.208	0.147	0.361
Ctrl. mean: fut. chrg.	0.094	0.193	0.144	0.332	0.081	0.157	0.118	0.279
Ctrl. comp. mean: fut. conv.	0.122	0.194	0.136	0.372	0.062	0.171	0.113	0.300
Ctrl. mean: fut. conv.	0.082	0.173	0.130	0.305	0.073	0.144	0.109	0.259
Ctrl. comp. mean: fut. incar.	0.066	0.070	0.062	0.192	0.034	0.098	0.059	0.185
Ctrl. mean: fut. incar.	0.059	0.127	0.091	0.232	0.050	0.102	0.076	0.192
Observations	104,224	104,224	104,224	104,224	79,159	79,159	79,159	79,159

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 four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). 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 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: 2SLS estimates for individuals from zip codes above and below median poverty level

	Above median poverty zip				Below median poverty zip			
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Conviction								
Fut. charge	0.173*	0.194	0.054	0.362**	0.016	-0.001	0.040	0.049
	(0.097)	(0.139)	(0.131)	(0.172)	(0.055)	(0.100)	(0.108)	(0.123)
Fut. conviction	0.166*	0.253*	0.005	0.394**	0.067	-0.012	0.041	0.116
	(0.087)	(0.132)	(0.124)	(0.158)	(0.051)	(0.095)	(0.103)	(0.120)
Fut. incarceration	0.112	0.188	-0.102	0.307**	0.061	-0.068	0.003	0.036
	(0.068)	(0.120)	(0.105)	(0.150)	(0.048)	(0.080)	(0.086)	(0.102)
Ctrl. comp. mean: fut. chrg.	0.165	0.319	0.188	0.471	0.128	0.252	0.210	0.440
Ctrl. mean: fut. chrg.	0.110	0.204	0.151	0.350	0.078	0.150	0.115	0.268
Ctrl. comp. mean: fut. conv.	0.144	0.284	0.182	0.438	0.115	0.230	0.202	0.416
Ctrl. mean: fut. conv.	0.091	0.177	0.133	0.316	0.067	0.133	0.102	0.242
Ctrl. comp. mean: fut. incar.	0.145	0.302	0.252	0.516	0.106	0.243	0.244	0.459
Ctrl. mean: fut. incar.	0.065	0.131	0.096	0.241	0.046	0.097	0.075	0.183
Panel B: Incarceration								
Fut. charge	-0.101**	0.004	0.077	0.006	-0.068*	0.041	-0.009	-0.015
	(0.046)	(0.073)	(0.066)	(0.086)	(0.042)	(0.067)	(0.060)	(0.080)
Fut. conviction	-0.093**	-0.026	0.101	-0.005	-0.094**	0.024	-0.002	-0.059
	(0.045)	(0.071)	(0.063)	(0.081)	(0.040)	(0.067)	(0.059)	(0.080)
Fut. incarceration	-0.056	-0.022	0.149***	0.038	-0.052	0.079	0.006	0.006
	(0.036)	(0.062)	(0.057)	(0.076)	(0.035)	(0.059)	(0.052)	(0.072)
Ctrl. comp. mean: fut. chrg.	0.145	0.194	0.137	0.364	0.094	0.205	0.170	0.374
Ctrl. mean: fut. chrg.	0.101	0.201	0.153	0.352	0.084	0.169	0.124	0.295
Ctrl. comp. mean: fut. conv.	0.099	0.155	0.094	0.279	0.066	0.181	0.139	0.330
Ctrl. mean: fut. conv.	0.088	0.180	0.138	0.324	0.075	0.155	0.114	0.274
Ctrl. comp. mean: fut. incar.	0.056	0.083	0.055	0.180	0.031	0.067	0.051	0.157
Ctrl. mean: fut. incar.	0.063	0.133	0.099	0.248	0.053	0.110	0.079	0.202
Observations	73,473	73,473	73,473	73,473	73,533	73,533	73,533	73,533

Note: The first panel of this table shows 2SLS estimates of the impact of conviction on future recidivism for individuals 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 four recidivism time ranges (1 year, 2–4 years, 5–7 years, and 1–7 years). 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 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.9: Alternative definitions of recidivism

	Noncarceral conviction				Incarceration			
	Year 1	Years 2-4	Years 5-7	Years 1-7	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Total number of charges and convictions								
Fut. charge	0.200	0.289	0.071	0.559	-0.154	-0.034	0.113	-0.074
	(0.202)	(0.414)	(0.302)	(0.577)	(0.114)	(0.215)	(0.167)	(0.313)
Fut. conviction	0.285	0.865**	0.327	1.477***	-0.158	-0.235	0.146	-0.246
	(0.183)	(0.356)	(0.256)	(0.496)	(0.103)	(0.202)	(0.152)	(0.283)
Ctrl Mean: fut. charge	0.222	0.499	0.362	1.083	0.213	0.484	0.353	1.050
Ctrl Mean: fut. conv.	0.000	0.000	0.000	0.000	0.209	0.473	0.344	1.025
Panel B: Total number of charge-events and conviction-events								
Fut. charge	0.101*	0.101	-0.088	0.114	-0.140***	-0.059	0.071	-0.128
	(0.061)	(0.129)	(0.121)	(0.212)	(0.039)	(0.078)	(0.061)	(0.117)
Fut. conviction	0.199***	0.358***	0.085	0.641***	-0.136***	-0.084	0.082	-0.138
	(0.056)	(0.115)	(0.101)	(0.193)	(0.036)	(0.075)	(0.061)	(0.113)
Ctrl Mean: fut. charge	0.115	0.251	0.183	0.549	0.108	0.247	0.184	0.540
Ctrl Mean: fut. conv.	0.000	0.000	0.000	0.000	0.106	0.241	0.179	0.526
Panel C: Recidivism with only 1 charge/conviction								
Fut. charge	0.047	-0.052	-0.009	0.045	-0.047*	0.060*	0.021	0.017
	(0.038)	(0.055)	(0.062)	(0.076)	(0.025)	(0.036)	(0.032)	(0.045)
Fut. conviction	0.076**	0.069	0.061	0.183***	-0.035	0.050	0.029	0.023
	(0.034)	(0.052)	(0.051)	(0.070)	(0.022)	(0.035)	(0.031)	(0.045)
Ctrl Mean: fut. charge	0.063	0.122	0.092	0.213	0.060	0.119	0.089	0.208
Ctrl Mean: fut. conv.	0.000	0.000	0.000	0.000	0.059	0.116	0.086	0.203
Panel D: Recidivism with 2 to 3 charges/convictions								
Fut. charge	0.051**	0.061	0.053	0.126***	-0.038**	-0.063***	-0.011	-0.080***
	(0.021)	(0.037)	(0.034)	(0.045)	(0.015)	(0.024)	(0.022)	(0.031)
Fut. conviction	0.068***	0.110***	0.074**	0.193***	-0.039***	-0.062***	-0.012	-0.089***
	(0.020)	(0.036)	(0.032)	(0.043)	(0.014)	(0.022)	(0.021)	(0.029)
Ctrl Mean: fut. charge	0.019	0.037	0.028	0.064	0.021	0.042	0.032	0.074
Ctrl Mean: fut. conv.	0.000	0.000	0.000	0.000	0.020	0.041	0.031	0.072
Panel E: Recidivism with 4 or more charges/convictions								
Fut. charge	0.004	0.067***	0.028	0.066**	-0.018**	-0.015	-0.007	-0.016
	(0.014)	(0.020)	(0.020)	(0.028)	(0.008)	(0.013)	(0.012)	(0.017)
Fut. conviction	0.017	0.086***	0.037*	0.101***	-0.016**	-0.017	-0.004	-0.018
	(0.013)	(0.019)	(0.019)	(0.027)	(0.007)	(0.013)	(0.011)	(0.016)
Ctrl Mean: fut. charge	0.005	0.011	0.008	0.019	0.006	0.013	0.010	0.023
Ctrl Mean: fut. conv.	0.000	0.000	0.000	0.000	0.006	0.013	0.010	0.022
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table shows 2SLS estimates of the impact of noncarceral conviction (Columns 1-4) and incarceration (Columns 5-8) on alternative definitions of recidivism. The first panel defines recidivism as the total number of future charges and convictions. The second panel is similar except recidivism is aggregated to the case level, defined by date. For instance, if a person receives two future charges on one date and another future charge an a separate date, this would count as three future charges in Panel A and two future charge-events on Panel B. Conviction-events are calculated similarly. Panels C-E show recidivism defined as only 1 charge/conviction, 2-3 charges/convictions, or 4 or more charges/convictions. Recidivism is measured from the date of sentencing using the time windows shown at top. 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 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.10: Testing for treatment-specificity

	Prior Conviction	Female	Black	Misdemeanors	Assualt	Burglary	Drugs	Fraud	Kidnapping	Larceny	Misc	Murder	Robbery	Sexual Assualt
Panel A: UPM($Z_c \mid Z_i$) – ordered model														
Conviction stringency (Z_c)	0.13***	-0.10**	0.0014	0.084**	-0.087*	0.036	0.026	0.055	-0.0020	0.047	-0.032**	-0.054*	0.022	0.0070
	(0.047)	(0.046)	(0.054)	(0.036)	(0.049)	(0.032)	(0.058)	(0.034)	(0.017)	(0.052)	(0.015)	(0.030)	(0.022)	(0.040)
Mean Dep. Var.	0.230	0.176	0.583	0.097	0.185	0.075	0.299	0.097	0.020	0.260	0.014	0.059	0.033	0.112
N	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692	153,692
Panel B: UPM($Z_i \mid Z_d$) – sequential and ordered model														
Incarceration stringency (Z_i)	-0.040	0.11	-0.059	0.0081	0.19**	-0.0060	0.14	-0.13**	0.016	-0.039	0.039*	0.028	0.048	-0.036
	(0.063)	(0.074)	(0.085)	(0.039)	(0.073)	(0.041)	(0.091)	(0.055)	(0.033)	(0.071)	(0.021)	(0.043)	(0.033)	(0.032)
Mean Dep. Var.	0.136	0.220	0.570	0.065	0.192	0.057	0.352	0.093	0.027	0.175	0.011	0.045	0.034	0.037
N	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589	28,589

Note: This table replicates the tests of the UPM assumption conducted in Table VI, but using individual covariates as the dependent variables rather than predicted recidivism. For Panel A, we restrict to the incarcerated sample and regress case characteristics on conviction stringency, controlling for incarceration stringency and court-by-time fixed effects. For Panel B, we restrict to the dismissed sample and regress case characteristics 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.

E.4   2SLS estimates with Empirical Bayes Shrinkage

We estimate judge stringency using leave-one-out means. To help ensure stringency measures are not too noisy, we restrict our analysis to judges who see at least 100 cases over the three-year windows we use to calculate stringency. We can further correct for potential measurement error using Empirical Bayes methods. Empirical Bayes was developed in the context of the teacher valued added literature (Chetty, Friedman, and Rockoff, 2014; Kane and Staiger, 2008), where the population distribution of teacher value added is typically assumed to be normally distributed, but measured with noise, also typically assumed to be normally distributed. This approach has also been applied to judge stringency measures in some papers (Arnold, Dobbie, and Hull, 2022; Norris, 2019), using standard Empirical Bayes shrinkage procedures (Morris, 1983).

We perform parametric Empirical Bayes, but we assume that judge stringencies are drawn from a Beta distribution, and the individual stringencies follow a Bernoulli distribution. We believe these parametric assumptions are better than assuming normality since judge stringencies are probabilities.

We take two approaches. The first assumes judge stringencies are drawn from a single Beta distribution, while the second assumes the Beta distribution varies by circuit and year.

Empirical Bayes with a single Beta prior.   First, we assume that judge stringencies are drawn from a $Beta(\alpha, \beta)$ distribution, and we estimate $\hat{\alpha}$ and $\hat{\beta}$ via maximum likelihood based on our sample of judge stringencies, which are calculated in three-year bins by judge, restricting to judges who handle at least 100 cases.[14] Let's consider noncarceral conviction stringency (the same derivations apply for incarceration or dismissal stringencies). Let C_j be the number of cases ending in a noncarceral conviction for judge j, and let N_j be the total number of cases they handle. Based on the estimated Beta prior, the posterior conviction stringency is given by

$$\frac{C_j + \alpha}{N_j + \alpha + \beta}.$$

We then construct the leave-one-out posterior stringency as

$$\frac{C_j - C_{j,i} + \alpha}{N_j - 1 + \alpha + \beta},$$

where i represents that particular case.

Figure E.1 plots our main stringency measures (x-axis) against the estimates Empirical Bayes estimates (y-axis). The measures are similar; they largely fall close to the 45 degree line.

Panel (a) of table E.1 reports our main first stage estimates; panel (b) reports the first-stage estimates using the shrunk stringency estimates. The results are very similar. The first-stage coefficients and F-statistics are slightly larger when we use the Empirical Bayes estimates. Panel (a) of table E.2 reproduces our main estimates, and panel (b) reports our 2SLS estimates for noncarceral conviviction using the Empirical Bayes stringencies. The

---

[14]To simplify, we use "judge" and j subscripts though, as in the rest of the paper, these are three-year rolling averages.

results are nearly identical. Table E.3 produces a similar table for incarceration, with similar conclusions.

Empirical Bayes with priors that vary by circuit-year. So far, we have used the same Beta prior for all judges. Here, we estimate priors that vary parametrically by circuit-year. We can express $\alpha = \gamma/\sigma$ and $\beta = (1-\gamma)/\sigma$, where γ is the average stringency and σ is the spread. We then estimate $\gamma_j = \gamma_0 + \gamma_{d,y}$, where $\gamma_{d,y}$ shifts the average stringency by circuit-year. We estimate this regression using a Bayesian Beta-Binomial regression, then with estimates of γ_0 and $\gamma_{d,y}$, we construct α_j and β_j for each judge-circuit-year. We construct the leave-one-out posterior stringency as

$$\frac{C_j - C_{j,i} + \alpha_j}{N_j - 1 + \alpha_j + \beta_j}.$$

This approach is very similar to the previous approach. The difference is that we now shrink judge stringencies towards the average stringency within circuit-year, rather than the overall average in our sample. This approach is appealing since it allows the prior distribution to vary by circuit-year, but requires estimating many more parameters to recover our empirical priors.

Analogous to panel (b), panel (c) of Table E.1 presents first-stage estimates using Empirical Bayes stringency with circuit-year priors. Here, we obtain first-stage coefficients that are closer to one, and larger F-statistics. A plausible interpretation is that this approach more effectively addresses measurement error in stringency measures.

Panel (c) of Table E.2 reports our main 2SLS estimates for noncarceral conviction using Empirical Bayes stringency with circuit-year priors. The results are very similar to our main specification, though estimates are somewhat smaller, particularly for the Years 1–7 time window, where estimates are 12.7% to 19.6% smaller and the estimate on future charge is statistically significant at the 0.1 rather than 0.05 level. Table E.3 produces similar results for the 2SLS estimates of incarceration with similar conclusions.

Overall, these results show that accounting for measurement error with either of the methods above does not qualitatively change our conclusions and does not lead to large quantitative differences.

Figure E.1: Leave-one-out stringency vs. leave-one-out Empirical Bayes stringency

(a) Dismissal

(b) Conviction

(c) Incarceration
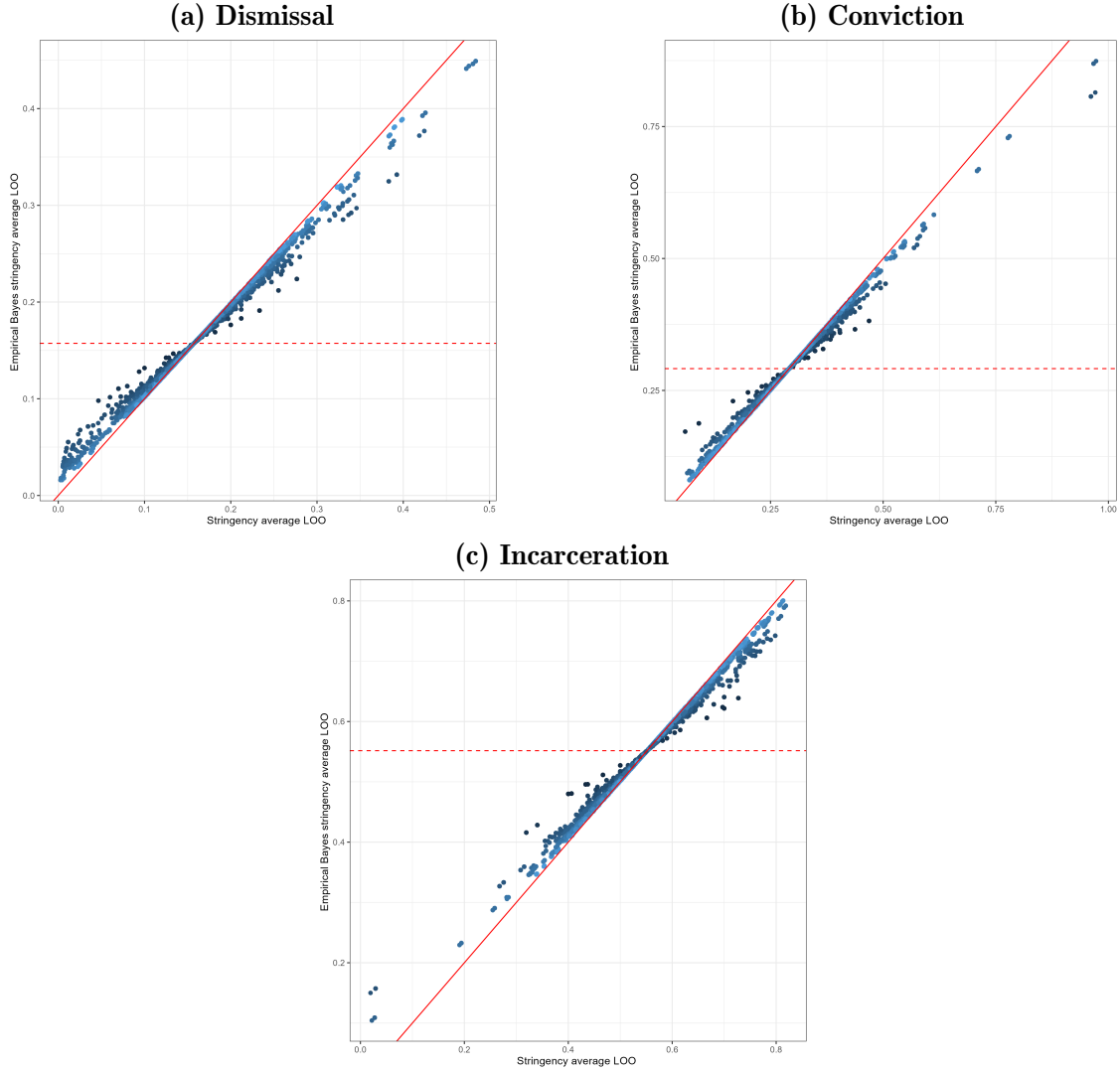
Notes: This figure compares the leave-one-out judge stringencies used in our main analysis to leave-one-out stringencies calculated via empirical Bayes with a single Beta prior. The lighter the blue points, the higher the total number of cases for judge j.

Table E.1: Relevance: first stage coefficients for the 2SLS analysis (Empirical Bayes shrinkage)

	Conviction			Incarceration		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: no shrinkage						
Conviction stringency	0.63***	0.60***	0.59***			
	(0.033)	(0.032)	(0.046)			
Incarceration stringency			-0.011	0.62***	0.59***	0.60***
			(0.041)	(0.033)	(0.032)	(0.035)
Dismissal stringency						0.032
						(0.051)
Controls	No	Yes	Yes	No	Yes	Yes
Mean dep. var.	0.298	0.298	0.298	0.546	0.546	0.546
F-stat	360.3	339.5	165.3	346.7	350.7	287.8
Observations	183,381	183,381	183,381	183,381	183,381	183,381
Panel B: single Beta-prior (Empirical Bayes leave-one-out)						
Conviction stringency	0.69***	0.65***	0.64***			
	(0.035)	(0.034)	(0.048)			
Incarceration stringency			-0.012	0.68***	0.65***	0.66***
			(0.044)	(0.035)	(0.033)	(0.037)
Dismissal stringency						0.029
						(0.055)
Controls	No	Yes	Yes	No	Yes	Yes
Mean dep. var.	0.298	0.298	0.298	0.546	0.546	0.546
F-stat	397.0	373.8	177.8	369.0	373.6	308.8
Observations	183,381	183,381	183,381	183,381	183,381	183,381
Panel C: priors varying by district-year (Beta Binomial leave-one-out)						
Conviction stringency	1.02***	0.97***	0.94***			
	(0.048)	(0.047)	(0.073)			
Incarceration stringency			-0.035	0.97***	0.94***	0.95***
			(0.066)	(0.050)	(0.047)	(0.051)
Dismissal stringency						0.042
						(0.094)
Controls	No	Yes	Yes	No	Yes	Yes
Mean dep. var.	0.298	0.298	0.298	0.546	0.546	0.546
F-stat	444.5	425.0	165.5	379.3	394.3	349.2
Observations	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table compares the coefficients on the instruments from the first stage of the 2SLS regressions in our main analysis (Panel A) with the coefficients derived using Empirical Bayes with a single Beta prior (Panel B) and with the coefficients derived using Empirical Bayes with priors that vary by circuit-year (Panel C).

Table E.2: Noncarceral conviction and recidivism (Empirical Bayes shrinkage)

	Year 1		Year 2-4		Year 5-7		Year 1-7	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Panel A: no shrinkage								
Fut. charge	-0.002	0.105**	0.004	0.087	0.006**	0.078	0.011***	0.235**
	(0.002)	(0.046)	(0.003)	(0.075)	(0.002)	(0.075)	(0.004)	(0.097)
Fut. conviction	0.001	0.135***	0.008***	0.115	0.007***	0.055	0.014***	0.300***
	(0.002)	(0.043)	(0.003)	(0.073)	(0.002)	(0.071)	(0.004)	(0.095)
Fut. incarceration	0.001	0.113***	0.006**	0.059	0.005**	-0.024	0.012***	0.215***
	(0.002)	(0.037)	(0.002)	(0.063)	(0.002)	(0.057)	(0.003)	(0.083)
Ctrl comp. mean: fut. chrg.	0.157	0.157	0.301	0.301	0.237	0.237	0.493	0.493
Ctrl mean: fut. chrg.	0.071	0.071	0.195	0.195	0.164	0.164	0.340	0.340
Ctrl comp. mean: fut. conv.	0.137	0.137	0.263	0.263	0.226	0.226	0.459	0.459
Ctrl comp. mean: fut. incar.	0.135	0.135	0.287	0.287	0.276	0.276	0.522	0.522
Ctrl mean: fut. incar.	0.048	0.048	0.137	0.137	0.112	0.112	0.250	0.250
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381
Panel B: single Beta prior (Empirical Bayes leave-one-out)								
Fut. charge	-0.002	0.109**	0.004	0.085	0.006**	0.077	0.011***	0.233**
	(0.002)	(0.046)	(0.003)	(0.072)	(0.002)	(0.073)	(0.004)	(0.094)
Fut. conviction	0.001	0.137***	0.008***	0.115*	0.007***	0.058	0.014***	0.299***
	(0.002)	(0.043)	(0.003)	(0.070)	(0.002)	(0.069)	(0.004)	(0.092)
Fut. incarceration	0.001	0.109***	0.006**	0.056	0.005**	-0.027	0.012***	0.207***
	(0.002)	(0.037)	(0.002)	(0.061)	(0.002)	(0.055)	(0.003)	(0.080)
Ctrl comp. mean: fut. chrg.	0.155	0.155	0.299	0.299	0.236	0.236	0.491	0.491
Ctrl mean: fut. chrg.	0.071	0.071	0.195	0.195	0.164	0.164	0.340	0.340
Ctrl comp. mean: fut. conv.	0.136	0.136	0.262	0.262	0.224	0.224	0.456	0.456
Ctrl mean: fut. conv.	0.063	0.063	0.176	0.176	0.149	0.149	0.313	0.313
Ctrl comp. mean: fut. incar.	0.133	0.133	0.284	0.284	0.273	0.273	0.517	0.517
Ctrl mean: fut. incar.	0.048	0.048	0.137	0.137	0.112	0.112	0.250	0.250
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381
Panel C: priors varying by district-year (Beta Binominal leave-one-out)								
Fut. charge	-0.002	0.102**	0.004	0.040	0.006**	0.090	0.011***	0.189*
	(0.002)	(0.047)	(0.003)	(0.075)	(0.002)	(0.075)	(0.004)	(0.097)
Fut. conviction	0.001	0.132***	0.008***	0.077	0.007***	0.069	0.014***	0.262***
	(0.002)	(0.045)	(0.003)	(0.072)	(0.002)	(0.072)	(0.004)	(0.097)
Fut. incarceration	0.001	0.113***	0.006**	0.020	0.005**	-0.024	0.012***	0.176**
	(0.002)	(0.039)	(0.002)	(0.063)	(0.002)	(0.058)	(0.003)	(0.084)
Ctrl comp. mean: fut. chrg.	0.131	0.131	0.268	0.268	0.213	0.213	0.447	0.447
Ctrl mean: fut. chrg.	0.071	0.071	0.195	0.195	0.164	0.164	0.340	0.340
Ctrl comp. mean: fut. conv.	0.115	0.115	0.238	0.238	0.200	0.200	0.416	0.416
Ctrl mean: fut. conv.	0.063	0.063	0.176	0.176	0.149	0.149	0.313	0.313
Ctrl comp. mean: fut. incar.	0.108	0.108	0.240	0.240	0.222	0.222	0.436	0.436
Ctrl mean: fut. incar.	0.048	0.048	0.137	0.137	0.112	0.112	0.250	0.250
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table compares the OLS and 2SLS regression estimates depicting the impact of noncarceral conviction on future recidivism in our main analysis (Panel A) with the estimates obtained using Empirical Bayes with a single Beta prior (Panel B) and with the coefficients derived using Empirical Bayes with priors that vary by circuit-year (Panel C).

Table E.3: Incarceration and recidivism (Empirical Bayes shrinkage)

	Year 1		Year 2-4		Year 5-7		Year 1-7	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Panel A: no shrinkage								
Fut. charge	-0.022***	-0.097***	0.013***	-0.016	0.025***	0.004	0.022***	-0.071
	(0.002)	(0.029)	(0.002)	(0.047)	(0.002)	(0.040)	(0.003)	(0.059)
Fut. conviction	-0.018***	-0.112***	0.013***	-0.038	0.023***	0.021	0.022***	-0.106*
	(0.001)	(0.029)	(0.002)	(0.047)	(0.002)	(0.039)	(0.003)	(0.058)
Fut. incarceration	-0.010***	-0.071***	0.017***	0.007	0.021***	0.052	0.027***	-0.029
	(0.001)	(0.024)	(0.002)	(0.041)	(0.002)	(0.032)	(0.003)	(0.051)
Ctrl comp. mean: fut. chrg.	0.122	0.122	0.199	0.199	0.147	0.147	0.370	0.370
Ctrl mean: fut. chrg.	0.088	0.088	0.173	0.173	0.131	0.131	0.303	0.303
Ctrl comp. mean: fut. conv.	0.084	0.084	0.168	0.168	0.113	0.113	0.311	0.311
Ctrl comp. mean: fut. incar.	0.043	0.043	0.071	0.071	0.051	0.051	0.166	0.166
Ctrl mean: fut. incar.	0.055	0.055	0.113	0.113	0.084	0.084	0.210	0.210
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381
Panel B: single Beta prior (Empirical Bayes leave-one-out)								
Fut. charge	-0.022***	-0.101***	0.013***	-0.016	0.025***	0.004	0.022***	-0.067
	(0.002)	(0.029)	(0.002)	(0.046)	(0.002)	(0.039)	(0.003)	(0.058)
Fut. conviction	-0.018***	-0.116***	0.013***	-0.038	0.023***	0.020	0.022***	-0.105*
	(0.001)	(0.028)	(0.002)	(0.046)	(0.002)	(0.038)	(0.003)	(0.056)
Fut. incarceration	-0.010***	-0.071***	0.017***	0.007	0.021***	0.050	0.027***	-0.030
	(0.001)	(0.024)	(0.002)	(0.040)	(0.002)	(0.032)	(0.003)	(0.050)
Ctrl comp. mean: fut. chrg.	0.124	0.124	0.200	0.200	0.147	0.147	0.373	0.373
Ctrl mean: fut. chrg.	0.088	0.088	0.173	0.173	0.131	0.131	0.303	0.303
Ctrl comp. mean: fut. conv.	0.087	0.087	0.170	0.170	0.113	0.113	0.314	0.314
Ctrl mean: fut. conv.	0.077	0.077	0.155	0.155	0.118	0.118	0.278	0.278
Ctrl comp. mean: fut. incar.	0.045	0.045	0.072	0.072	0.051	0.051	0.169	0.169
Ctrl mean: fut. incar.	0.055	0.055	0.113	0.113	0.084	0.084	0.210	0.210
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381
Panel C: priors varying by district-year (Beta Binominal leave-one-out)								
Fut. charge	-0.022***	-0.106***	0.013***	0.001	0.025***	-0.003	0.022***	-0.056
	(0.002)	(0.028)	(0.002)	(0.045)	(0.002)	(0.038)	(0.003)	(0.056)
Fut. conviction	-0.018***	-0.117***	0.013***	-0.023	0.023***	0.013	0.022***	-0.096*
	(0.001)	(0.028)	(0.002)	(0.045)	(0.002)	(0.037)	(0.003)	(0.054)
Fut. incarceration	-0.010***	-0.074***	0.017***	0.016	0.021***	0.044	0.027***	-0.027
	(0.001)	(0.023)	(0.002)	(0.039)	(0.002)	(0.030)	(0.003)	(0.048)
Ctrl comp. mean: fut. chrg.	0.112	0.112	0.193	0.193	0.142	0.142	0.353	0.353
Ctrl mean: fut. chrg.	0.088	0.088	0.173	0.173	0.131	0.131	0.303	0.303
Ctrl comp. mean: fut. conv.	0.081	0.081	0.165	0.165	0.115	0.115	0.302	0.302
Ctrl mean: fut. conv.	0.077	0.077	0.155	0.155	0.118	0.118	0.278	0.278
Ctrl comp. mean: fut. incar.	0.045	0.045	0.085	0.085	0.061	0.061	0.180	0.180
Ctrl mean: fut. incar.	0.055	0.055	0.113	0.113	0.084	0.084	0.210	0.210
Observations	183,381	183,381	183,381	183,381	183,381	183,381	183,381	183,381

Note: This table compares the OLS and 2SLS regression estimates depicting the impact of incarceration on future recidivism in our main analysis (Panel A) with the estimates obtained using Empirical Bayes with a single Beta prior (Panel B) and with the coefficients derived using Empirical Bayes with priors that vary by circuit-year (Panel C).

E.5 Calculating control means for compliers

To calculate control-group complier means, we follow Dahl, Kostøl, and Mogstad (2014) and Agan and Starr (2018). First we show how to derive control-group complier means for the simple case of binary treatment and a binary instrument. We then expand this to our setting.

In the simple case where $Z \in 0, 1$ and $D \in 0, 1$, we aim to calculate $E[Y(0) \mid D(1) > D(0)]$. Here $Y(0)$ is the potential outcome when $D = 0$, $D(1)$ is the potential treatment when $Z = 1$, and $D(0)$ is the potential treatment when $Z = 0$. Note that

$$\underbrace{E[Y \mid D = 0, Z = 0]}_{\text{data}} = \frac{\pi_c}{\pi_c + \pi_n} \underbrace{E[Y(0) \mid D(1) > D(0)]}_{\text{unknown}} + \frac{\pi_n}{\pi_c + \pi_n} E[Y(0) \mid D(1) = D(0) = 0],$$

where

$$\pi_n = \underbrace{Pr(D = 0 \mid Z = 1)}_{\text{data}}$$

$$\pi_a = \underbrace{Pr(D = 1 \mid Z = 0)}_{\text{data}}$$

$$\pi_c = 1 - \pi_n - \pi_a.$$

In the expression above, the terms with "data" below them can be calculated directly from the data. The term $E[Y(0) \mid D(1) = D(0) = 1] = E[Y \mid D = 0, Z = 1]$, where the right-hand term can also be calculated directly from the data. This leaves only one unknown term: $E[Y(0) \mid D(1) > D(0)]$, which is the term of interest. Re-arranging the equations and plugging in, we get:

$$E[Y(0) \mid D(1) > D(0)] = \frac{\pi_c + \pi_n}{\pi_c} E[Y \mid D = 0, Z = 0] - \frac{\pi_n}{\pi_c} E[Y \mid D = 0, Z = 1],$$

where all the terms on the right side of the equality can be estimated from the data.

Our setting differs from the simple setting above because we have a continuous instrument, and D can take on 3 values. We follow Dahl, Kostøl, and Mogstad (2014) and Agan and Starr (2018) in adapting the math above to the case with continuous instruments. We use code from the replication file of Agan and Starr (2018), which is adapted from Dahl, Kostøl, and Mogstad (2014). This adaptation involves calculating the minimum and maximum values of the instrument (z_{min} and z_{max}). Following the papers above, we can then adapt the equations to be

$$\underbrace{E[Y \mid D = 0, Z = z_{min}]}_{\text{data}} = \frac{\pi_c}{\pi_c + \pi_n} \underbrace{E[Y(0) \mid D(z_{max}) > D(z_{min})]}_{\text{unknown}}$$

$$+ \frac{\pi_n}{\pi_c + \pi_n} E[Y(0) \mid D(z_{max}) = D(z_{min}) = 0],$$

where

$$\pi_n = \underbrace{Pr(D = 0 \mid Z = z_{max})}_{\text{data}}$$

$$\pi_a = \underbrace{Pr(D = 1 \mid Z = z_{min}])}_{\text{data}}$$

$$\pi_c = \beta * (z_{max} - z_{min}),$$

and where β is from the regression of D on the instrument. Similar to the binary case, we have $E[Y(0) \mid D(z_{min}) = D(z_{max}) = 1] = E[Y \mid D = 0, Z = z_{max}]$. We use the first and 99th percentiles of the residualized instrument for z_{min} and z_{max}, respectively.

To address the fact that we consider multiple treatments, we include non-focal judge stringency as an additional control. For example, if D is the indicator for conviction, we use judge conviction stringency as the instrument, controlling for judge incarceration stringency. Under UPM and A1–A4 in the main paper, the only compliers will be those shifting from $T = d$ to $T = c$, capturing the margin-specific compliers of interest.

F ADDITIONAL DETAILS FOR MULTINOMIAL MODEL WITH HETEROGENEOUS EFFECTS

This appendix discusses how we apply Mountjoy (2022) in our setting. First we discuss the identification of treatment-specific thresholds from judge stringencies. Second, we describe the identification and estimation of margin-specific treatment effects. Then, we report additional empirical results.

F.1 Additional details on identification and estimation

We start with the identification of treatment-specific instruments, and then discuss how we adapt the identification and estimation strategies from Mountjoy (2022) to obtain the results in Section V.

The first step in our approach is to identify judges' choice-specific thresholds (π_c and π_i in the unordered multinomial choice model in Section *III.C.3* of the paper) based off of the shares of cases ending in each outcome, Z_c and Z_i. Berry, Gandhi, and Haile (2013) show that the inversion between shares and thresholds exists under weak assumptions. Specifically, 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 a specific option 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 assumption is satisfied if the support of the additive errors is $\mathbb{R}^K$, where K is the number of choices.

Next, Berry and Haile (2024) show that judge-specific thresholds can be identified without invoking identification at infinity arguments. Their argument assumes an index structure on the structural choice probability function where judges' latent preferences enter linearly into the index. Using this setup, their 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 each judge. In particular, identification requires three continuous covariates whose loadings do not vary across judges. This assumption is very similar to the standard monotonicity conditions invoked in applied judge IV papers, as judge stringencies typically enter linearly and are not estimated conditional on covariates (an exception is Mueller-Smith, 2015). Similarly, when covariates are included, they also typically enter linearly, with loadings that do not vary by judge. While this could be seen as a strong assumption, it is one we make for tractability and to allow us to focus on other aspects of the research design. 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. Kamat, Norris, and Pecenco (2024) provide an alternative approach that does not require covariates, but it uses a sequential model and recovers bounds rather than point estimates. As described in the paper and below, we make additional assumptions for tractability when estimating judge thresholds, and show that the results are broadly similar under a few different variations on these assumptions.

Next, we discuss how we adapt the identification and estimation strategies from Mountjoy (2022) to obtain the results in Section V. To begin, we state the "comparable compliers" assumption of Mountjoy (2022) in our notation:

A7. Comparable Compliers (CC)

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

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

This assumption says that $i \rightarrow c$ compliers from decreasing $\tilde{z}_i$ have the same potential outcome when convicted as $i \rightarrow c$ 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 two LATEs that are specific to two different margins as captured by the following expression:

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

We visualize this decomposition in Panel (c) of Figure III, 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 green).

Mountjoy (2022) shows that it is possible to recover the two margin-specific LATEs, as well as ω, by using variation in two treatment-specific instruments to construct the relevant expected potential outcomes for the two groups. His identification results directly apply once we have recovered treatment-specific instruments.

We also follow Mountjoy (2022) in estimation. For example, we assume the relevant conditional expectations are well approximated by a local linear regression centered around the chosen evaluation point of the instruments. These regressions include additive controls as specified in the notes of Table VII. 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 500 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.

F.2 Additional results

Tables F.1 and F.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.[15] While restrictive, this allows for a simple closed-form solution for constructing thresholds from shares, as explained in the main paper. The second set of results mirrors the mixed model reported in Table VII, but assumes the random effects follow an independent multivariate normal distribution. We calculate confidence intervals for all three approaches using 500 bootstrap samples.

Overall, the results in Tables F.1 and F.2 are similar to Table VII.

---

[15]In this model the thresholds are simply $\pi_c(z_c, z_i) = log(z_c) - log(1 - z_c - z_i)$ and $\pi_i(z_c, z_i) = log(z_i) - log(1 - z_c - z_i)$.

Table F.1: Margin-specific treatment effects: alternative approach (robustness, multinomial logit)

	Simple log-ratio			
	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Noncarceral conviction vs dismissal (C vs D)				
Felony charge:	0.067	0.171*	0.187***	0.262**
	[-0.038,0.186]	[-0.012,0.387]	[0.050,0.383]	[0.041,0.512]
	{0.077}	{0.173}	{0.136}	{0.343}
Felony conviction:	0.090*	0.215***	0.156**	0.341***
	[-0.012,0.205]	[0.053,0.402]	[0.035,0.328]	[0.135,0.644]
	{0.068}	{0.132}	{0.146}	{0.282}
Felony incarceration:	0.056	0.149**	0.074	0.212**
	[-0.035,0.150]	[0.009,0.307]	[-0.049,0.223]	[0.019,0.444]
	{0.070}	{0.107}	{0.123}	{0.297}
Panel B: Incarceration vs noncarceral conviction (I vs C)				
Felony charge:	-0.045**	0.034	0.002	-0.039
	[-0.080,-0.009]	[-0.030,0.104]	[-0.074,0.063]	[-0.150,0.074]
	{0.084}	{0.178}	{0.138}	{0.334}
Felony conviction:	-0.037*	0.029	0.019	-0.030
	[-0.066,0.002]	[-0.034,0.099]	[-0.045,0.090]	[-0.139,0.081]
	{0.074}	{0.163}	{0.120}	{0.306}
Felony incarceration:	-0.013	0.043	0.014	-0.040
	[-0.043,0.016]	[-0.023,0.099]	[-0.041,0.074]	[-0.152,0.051]
	{0.054}	{0.109}	{0.099}	{0.241}
Controls	Yes	Yes	Yes	Yes

Note: This table reports margin-specific estimates of the impact of noncarceral conviction vs dismissal (Panel A) and incarceration vs noncarceral conviction (Panel B) using an unordered multinomial model based on the methodology developed in Mountjoy (2022). We follow the methods described in Section V, except that here judge-specific latent preferences are calculated under the assumption that case outcomes are determined by a multinomial logit. The curly brackets report control-group complier means. 95% confidence intervals are reported in brackets and are based on 500 bootstrap samples. Stars denote * p< 0.10, ** p< 0.05, *** p< 0.01 based on the 90%, 95%, and 99% confidence intervals.

Table F.2: Margin-specific treatment effects: alternative approach (robustness, independent mixed logit)

	Mixed logit with independent normal random effects			
	Year 1	Years 2-4	Years 5-7	Years 1-7
Panel A: Noncarceral conviction vs dismissal (C vs D)				
Felony charge:	0.060	0.154**	0.104**	0.182**
	[-0.012,0.124]	[0.059,0.272]	[0.005,0.206]	[0.038,0.328]
	{0.067}	{0.149}	{0.124}	{0.297}
Felony conviction:	0.070*	0.169***	0.085	0.226***
	[-0.001,0.141]	[0.058,0.278]	[-0.015,0.191]	[0.083,0.374]
	{0.057}	{0.125}	{0.125}	{0.252}
Felony incarceration:	0.048*	0.116***	0.048	0.150**
	[-0.008,0.105]	[0.029,0.208]	[-0.046,0.131]	[0.009,0.303]
	{0.051}	{0.102}	{0.094}	{0.232}
Panel B: Incarceration vs noncarceral conviction (I vs C)				
Felony charge:	-0.052***	0.023	-0.013	-0.078*
	[-0.083,-0.021]	[-0.031,0.079]	[-0.070,0.046]	[-0.171,0.014]
	{0.088}	{0.174}	{0.146}	{0.349}
Felony conviction:	-0.043***	0.016	-0.001	-0.072*
	[-0.077,-0.011]	[-0.041,0.072]	[-0.055,0.052]	[-0.167,0.008]
	{0.078}	{0.162}	{0.128}	{0.324}
Felony incarceration:	-0.018	0.033	0.002	-0.070
	[-0.043,0.007]	[-0.011,0.086]	[-0.045,0.050]	[-0.144,0.016]
	{0.056}	{0.106}	{0.098}	{0.251}
Controls	Yes	Yes	Yes	Yes

Note: This table reports margin-specific estimates of the impact of noncarceral conviction vs dismissal (Panel A) and incarceration vs noncarceral conviction (Panel B) using an unordered multinomial model based on the methodology developed in Mountjoy (2022). We follow the methods described in Section V, except that here judge-specific latent preferences are calculated under the assumption that the intercepts include a random effect that is an uncorrelated multivariate normal. The curly brackets report control-group complier means. 95% confidence intervals are reported in brackets and are based on 500 bootstrap samples. Stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on the 90%, 95%, and 99% confidence intervals.

G   IMPACTS OF INCARCERATION: ADDITIONAL EVIDENCE FROM SENTENCING GUIDELINES

In this Appendix, we provide supporting evidence on the effects of incarceration, exploiting an independent source of variation: discontinuous changes in recommended sentences in the Virginia sentencing guidelines. Although judges have the final say over sentencing in Virginia, each person convicted of a felony gets a guidelines-recommended sentence that is calculated using a series of worksheets. Sentence recommendations change discontinuously at some scores. 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 individuals who had never previously been incarcerated.

G.1   Empirical setup

Sample and data.   For these analyses, we focus on people who were convicted of a felony in Virginia and use data from the Virginia Criminal Sentencing Commission (VCSC). See Appendix B for more details on the data and sample construction, and Table G.1 for summary statistics on our sample.

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 G.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 score assigned to 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. We include an example of Worksheet A in Appendix G.7; the other worksheets follow a similar organization.

Empirical approach.   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 assumption 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 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 this score is less salient. Third, from the legislator's standpoint, the goal of these worksheets was to reduce unjustified disparities. Therefore, it seems unlikely that the sharp sentencing discontinuities observed at the cutoff in the incarceration-length score were created on purpose. In Section G.4 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"—designed specifically for regression discontinuity with a discrete running variable. As in other RD settings, we want to estimate a function of the form:

$$Y_{i,s} = \beta \cdot \mathbb{1}(s \geq 0) + f(s) \cdot \mathbb{1}(s \geq 0) + g(s) \cdot \mathbb{1}(s < 0) + \epsilon, \qquad (12)$$

where $Y_{i,s}$ is the outcome of the person in case i having obtained a sentencing score of s.[16] Our main coefficient of interest is β. The challenge is to estimate the form of $f(\cdot)$ and $g(\cdot)$, 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 $f(\cdot)$ and $g(\cdot)$ close to the cutoff, based on the functions' behavior further from the cutoff, to estimate the magnitude of plausible deviations from the linear estimation. We need to choose a parameter K that is the upper bound of the absolute value of the second derivative of the conditional expectation function. That parameter tells us how quickly the functions $f(\cdot)$ and $g(\cdot)$ 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

---

[16]The sentencing score is either the incarceration-length score or the probation/jail score.

behavior at other points.

To choose K, we follow the approach developed by Imbens and Wager (2019) and implemented by Goldsmith-Pinkham, Pinkovskiy, and Wallace (2023). 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.[17] 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 $f(\cdot)$ at the cutoff to be double 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. To keep bandwidths constant across outcomes and time periods, we use a bandwidth of 5 in all specifications.

G.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 G.2 show the regression discontinuity results, and Appendix Figure G.4 graphs these results. Scoring above the threshold generates large (44 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.[18]

By comparing people on either side of the threshold, we can estimate the causal effect on recidivism of going from a sentence of approximately four months to approximately one year. Columns 3–9 of Table G.2 present outcomes for various time periods, from year 1 to years 1–7 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, 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 1–7 year cumulative measure, we can reject any increase in new felony charges larger than 0.2 percentage points over the control group mean of 41%.

G.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. The impacts of incarceration may accrue rapidly in the first several months. For example, a few months in jail might lead a person to lose their job, as documented in the pretrial detention context (Dobbie, Goldin, and Yang, 2018), or to experience ruptures in their family lives. 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 G.3 show that scoring above the threshold translates into a 43 ppt increase in the likelihood of receiving a carceral sentence,

---

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

[18]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.

and the average sentence length increases by 0.76 months (Figure G.5 graphs this extensive margin). Estimates from the probation/jail sample therefore capture the effect of a short jail sentence relative to probation only.[19] Columns 3–5 of Panel A of Table G.3 present results for recidivism. Given that sentences around the cutoff are so short, we look at short-term results using the six months after sentencing, and longer-term results looking 2–3 years after sentencing. We find no evidence of a short-term incapacitation effect—perhaps because the difference in sentences is only about a month.[20] We find no evidence of longer-term effects. In our 1–3 years cumulative measure we can reject anything larger than a 0.7 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. 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.[21] This substantially decreases our sample size, particularly since data on pretrial detention is only available after 2010. As seen in Panel B of Table G.3, 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 from exploiting quasi-random assignment of cases to judges: we find short-term decreases in criminal justice contact, consistent with incapacitation, but we do not detect any longer-term impacts of exposure to incarceration. Table E.2 Panel B and Table G.6 present complier characteristics for the IV analyses, and characteristics of defendants who score just above or just below the relevant cutoffs. These groups are similar, but there are some small differences. For example, marginal defendants in the RD analysis are more likely to have been convicted for a drug crime compared to the IV compliers—especially for the extensive margin analyses.

G.4 Balance and marginal cases

Balance tests. Figure G.2 (G.3) and Table G.4 (G.5) present balance tests for the intensive margin experiment based on Worksheet A (extensive margin experiment, Worksheet B). We first perform 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 to the cutoff, moving them just below the cutoff and avoiding a longer recommended sentence. If defense attorneys are trying to push their clients to the left of the cutoff, more charges may be dropped just before the cutoff because some of the points are linked to the number of offenses for which a person is convicted – a pattern we do not observe in our data. We also

---

[19]Short sentences such as those experienced right above the cutoff are not atypical. For example, in Pennsylvania, individuals released from jail had spent an average of 2.4 months incarcerated after sentencing (PASC, 2013).

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

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

look at measures of defendant poverty, which can affect the quality of representation (Agan, Freedman, and Owens, 2021).[22] 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 2.3 percentage points more likely to have their case resolved by plea if they fall just before the cutoff (Panel B of table G.4). One interpretation is that 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 2.3 percentage point increase in the trial rate would lead to losing 0.23% of the sample right above the threshold. Given how small the difference in conviction is at the threshold, and the fact that we see no detectable differences in observable characteristics, we think this difference is unlikely to affect our research design. 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 G.2 (a) and G.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. However, we also see excess mass in other places of the distribution, 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 cases. Appendix Table G.6 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: 85% of individuals had been incarcerated in the past, compared to 64% for the sample overall. Marginal cases are similar across 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.

---

[22]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.

G.5 Appendix figures: RD analyses

Figure G.1: Flowchart of felony sentencing determination in Virginia
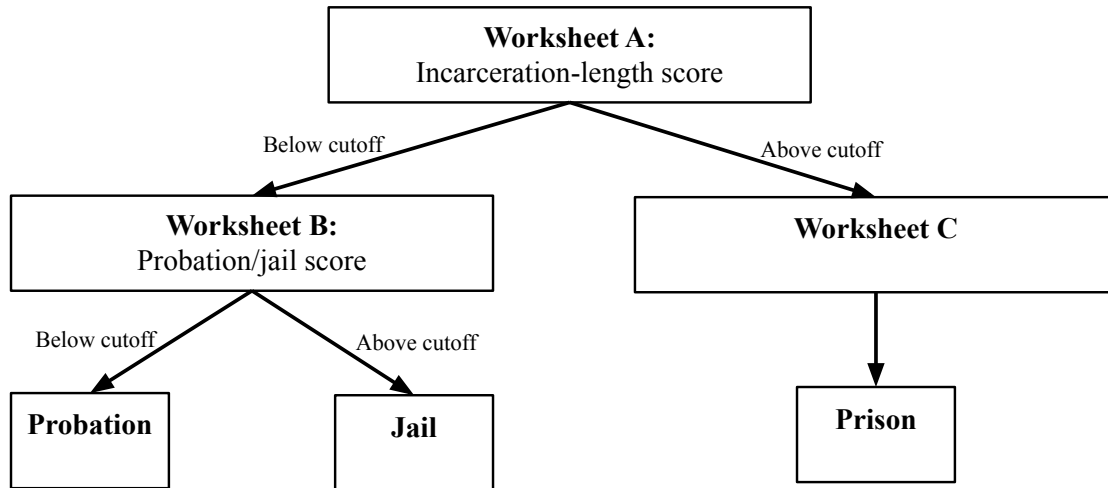
Note: This figure presents a flowchart describing the sentencing process in Virginia after a felony conviction, and how and when different Worksheets are used.

Figure G.2: Balance tests—incarceration-length sample
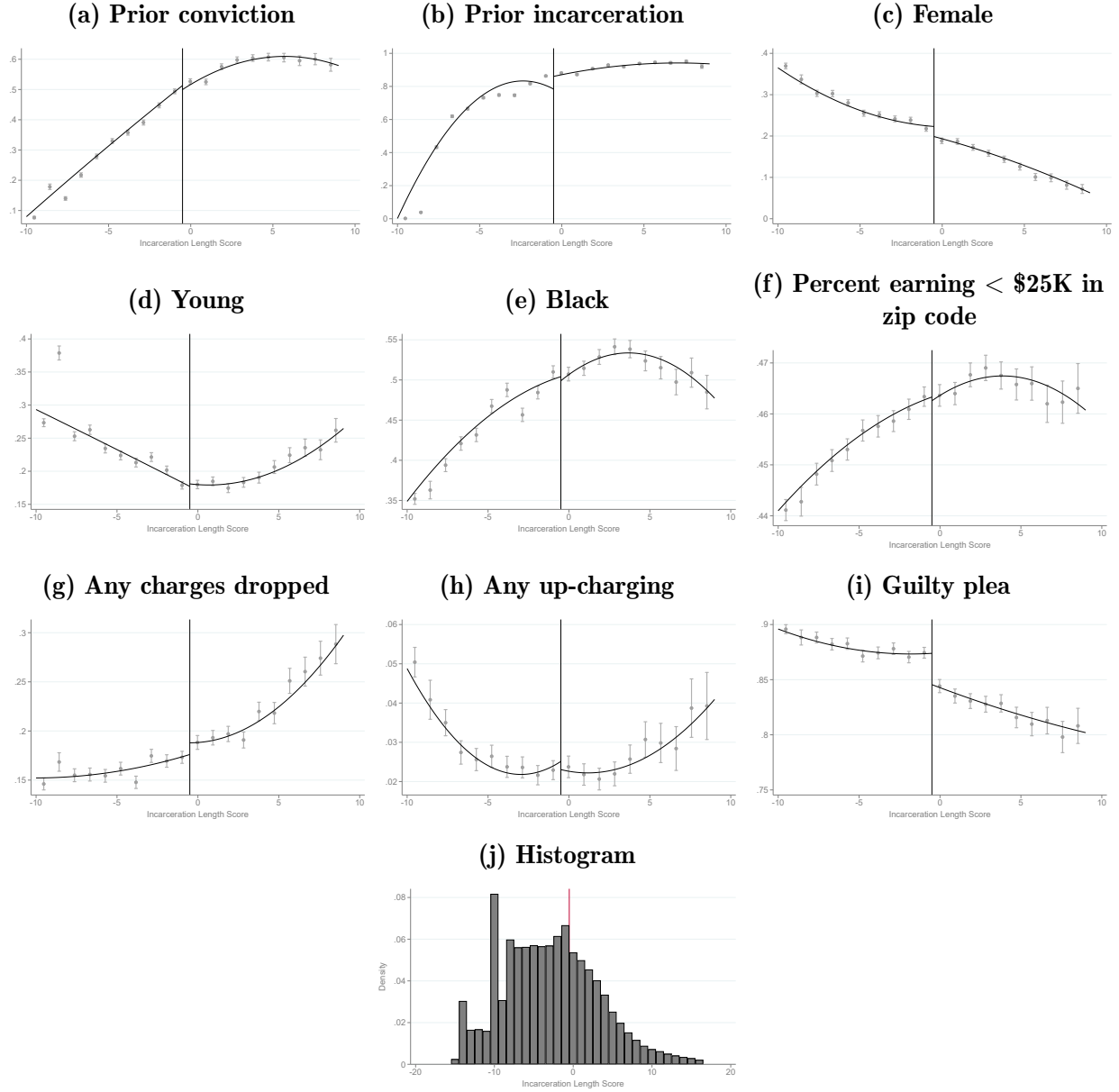
(a) Prior conviction

(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 various demographic variables and case characteristics around the cutoff in the incarceration-length score. 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 G.3: Balance tests—probation/jail sample
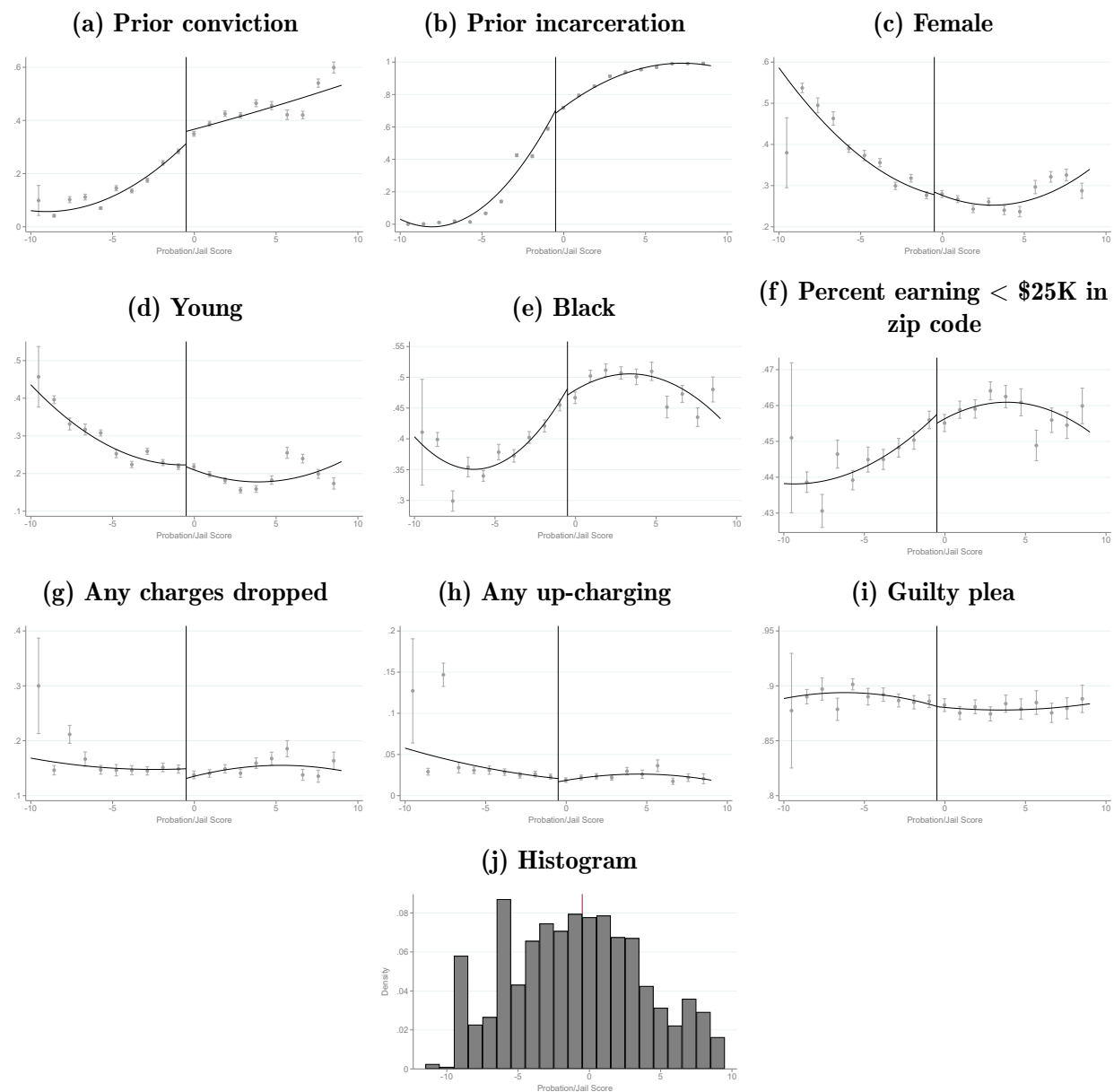
(a) Prior conviction

(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 various demographic variables and case characteristics around the cutoff in the probation/jail score. 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 G.4: RD first stage and outcome graphs—incarceration-length sample

(a) Incarcerated for at least 1 year
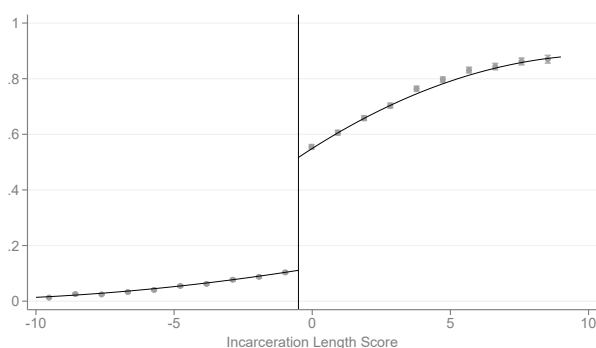
(b) Sentence in months
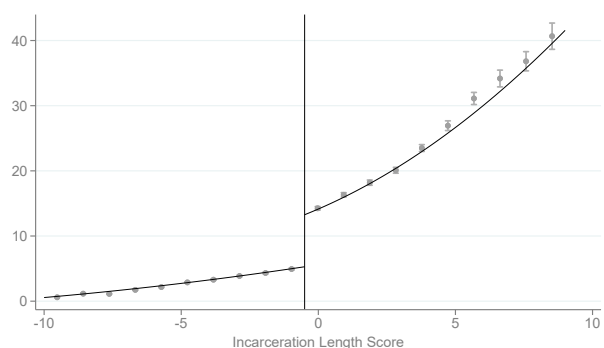
(c) Incarcerated for at least five years
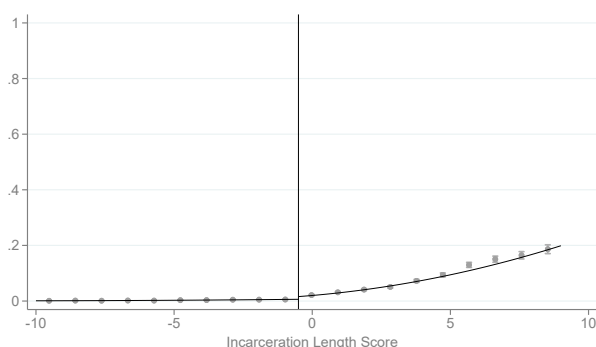
(d) Sentencing distributions for those right above/below the cutoff
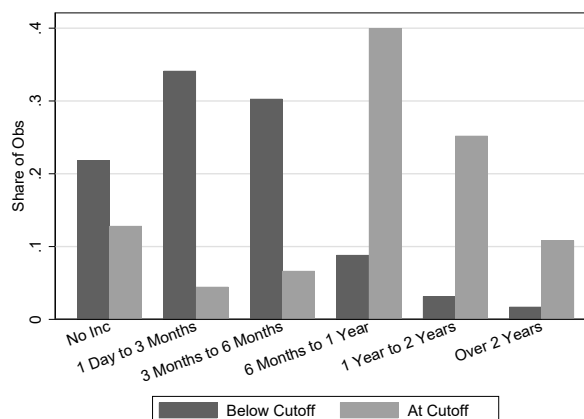
(e) New charge within 1 year of sentencing
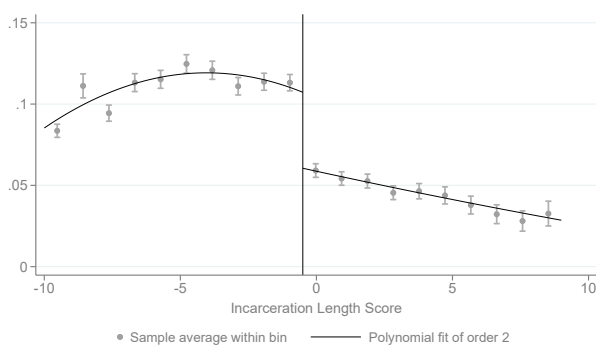
(f) New charge in years 5-7 after sentencing
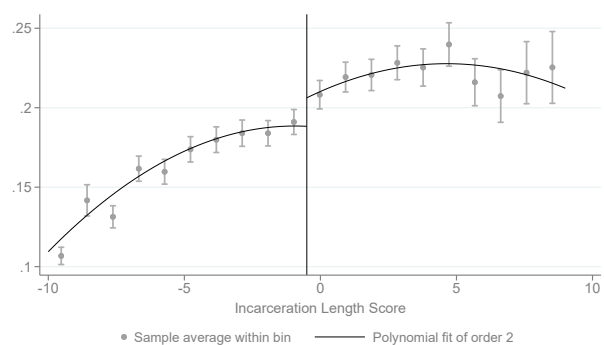
Note: Panel (a) shows the discontinuity in being incarcerated for at least one year around the cutoff in the incarceration-length score. Panel (b) shows the same plot for months sentenced, and panel (c) shows the same plot for being sentenced to at least five years. Panel (d) shows the distribution of sentence lengths for those just above and just below the cutoff. Panel (e) shows the plot for recidivism—defined as a binary variable for having at least one new charge one year post-sentencing—and panel (f) shows recidivism within 5–7 years post-sentencing.

Figure G.5: RD First stage and outcome graphs—probation/jail score

(a) Incarcerated at all

(b) Sentence in months

(c) Incarcerated for at least one year

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

(e) New charge within 6 months of sentencing

(f) New charge in years 2-3 after sentencing
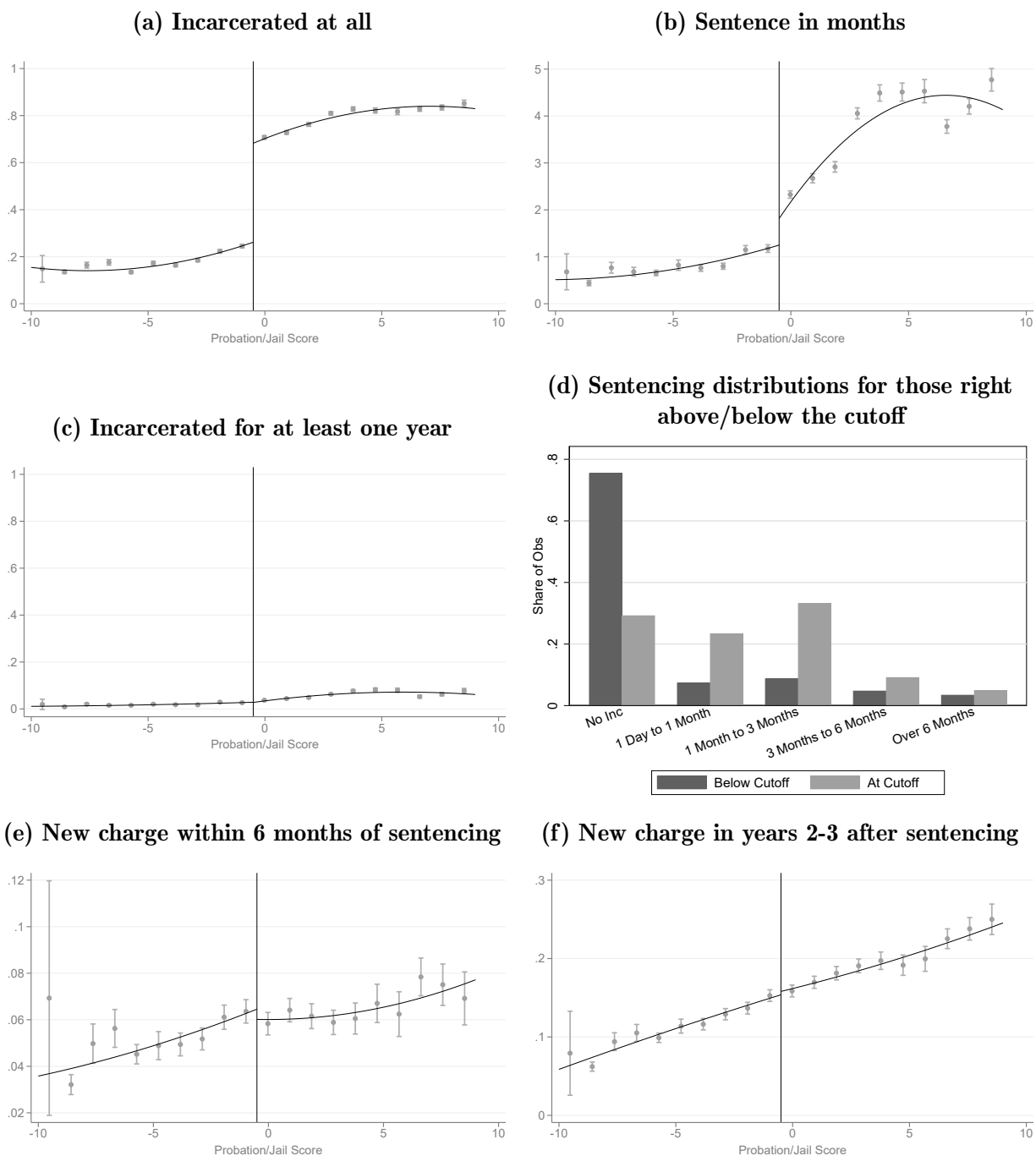
Note: Panel (a) shows the discontinuity in being incarcerated at all. Panel (b) shows the same plot for months sentenced and panel (c) shows the same plot for being sentenced to at least one year. Panel (d) shows the distribution of sentence lengths for those just above and just below the cutoff. Panel (e) shows the 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.

G.6 Appendix tables: RD analyses

Table G.1: Summary statistics: RD sample

	Incarceration length worksheet	Probation/jail worksheet
	mean	mean
Offenses		
Assault	0.05	0.00
Burglary	0.11	0.00
Drug	0.41	0.57
Larceny	0.35	0.42
Miscellaneous	0.02	0.01
Robbery	0.02	0.00
Sexual assault	0.03	0.00
Defendant characteristics		
Black	0.50	0.45
Female	0.23	0.32
Under 23	0.26	0.24
% of ppl in zip earning <25K	0.46	0.45
Recommended for prison	0.34	0.00
Prior incarceration	0.63	0.54
Prior circuit crt. felony convic.	0.33	0.27
Sentencing		
Carceral sentence	0.61	0.47
Jail sentence	0.34	0.45
Prison sentence	0.28	0.04
Sentence >= 5 years	0.04	0.00
Months of sentence	10.50	2.15
Observations	151,751	115,266

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

Table G.2: 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) 1-7 years
Treatment	0.441	8.474	-0.052	-0.009	0.015	-0.023
	[0.422,0.460]	[7.898,9.051]	[-0.065,-0.039]	[-0.028,0.010]	[-0.004,0.034]	[-0.048,0.002]
N	81,440	81,440	81,440	81,440	81,440	81,440
Control mean	0.08	4.00	0.12	0.23	0.18	0.41

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 1–2) and recidivism (columns 3–6). 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, during which incarceration rates across treatment and control are equal. It also shows cumulative time windows of 1–7 years to compare to our IV estimates. Below the estimates, we present confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Appendix G for a discussion of parameter choices.

Table G.3: Incarceration and recidivism: RD estimates for the extensive margin

	Sentence		Recidivism		
	(1) **Any Incar**	(2) **Months**	(3) **6 months**	(4) **1-3 years**	(5) **2-3 years**
Panel A: probation/jail sample					
Treatment:	0.428	0.756	-0.007	-0.006	-0.003
	[0.391,0.465]	[0.524,0.988]	[-0.014,0.001]	[-0.020,0.007]	[-0.015,0.009]
N	80,286	80,286	80,286	80,286	80,286
Control mean	0.21	0.98	0.06	0.21	0.13
Panel B: no prior incar. probation/jail sample					
Treatment:	0.423	0.922	0.013	0.022	0.001
	[0.342,0.505]	[0.289,1.555]	[-0.015,0.042]	[-0.037,0.080]	[-0.043,0.045]
N	7,875	7,875	7,875	7,875	7,875
Control mean	0.18	0.80	0.05	0.20	0.12

Note: This table first shows the RD estimates of how the cutoff affects sentences (probability of getting a carceral sentence and sentence length (columns 1–2) and recidivism (columns 3–5). We measure recidivism as 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. It also shows cumulative 1–3 year estimates to compare more closely to our IV results. The third is years 2–3, during which incarceration rates across treatment and control are equal. 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 confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff. See Appendix G for a discussion of parameter choices.

Table G.4: 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	0.006	-0.004	-0.016	0.002	-0.015
	[-0.020,0.012]	[-0.011,0.023]	[-0.172,0.165]	[-0.033,0.002]	[-0.038,0.042]	[-0.043,0.012]
N	81,440	81,265	81,440	81,440	81,022	81,440
Control mean	0.80	0.35	0.77	0.22	0.22	0.53
	Plea	Dropped Chrg.	Upgrade Chrg.	Zip <10K	Zip <25K	
Panel B: income & legal actor balance						
RD Estimate:	-0.023	0.006	-0.001	-0.000	-0.000	
	[-0.040,-0.005]	[-0.014,0.026]	[-0.009,0.007]	[-0.003,0.002]	[-0.006,0.005]	
N	81,440	76,448	76,448	58,901	58,901	
Control mean	0.85	0.16	0.02	0.19	0.47	

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 confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff.

Table G.5: 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:	-0.007	0.033	0.040	0.017	0.007	-0.001
	[-0.034,0.021]	[-0.014,0.079]	[-0.105,0.184]	[-0.025,0.058]	[-0.032,0.046]	[-0.051,0.049]
N	80,286	80,096	80,286	80,286	79,948	80,284
Control mean	0.79	0.20	0.36	0.31	0.24	0.43
	Plea	Dropped Chrg.	Upgrade Chrg.	Zip <10K	Zip <25K	
Panel B: zip income & legal actor balance						
RD Estimate:	-0.001	-0.012	-0.003	-0.001	-0.003	
	[-0.015,0.013]	[-0.026,0.002]	[-0.008,0.003]	[-0.003,0.001]	[-0.008,0.003]	
N	80,286	72,566	72,566	57,010	57,010	
Control mean	0.88	0.14	0.02	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 confidence intervals obtained following Kolesár and Rothe (2018). Our estimations are for a bandwidth of 5 above and below the cutoff.

Table G.6: Marginal cases in the RD study

	Incarceration length worksheet		Probation/jail worksheet	
	$Pr(X = x)$	$Pr(X = x\|Marginal)$	$Pr(X = x)$	$Pr(X = x\|Marginal)$
Prior Conviction	0.636	0.852	0.521	0.564
	(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	0.650	0.871	0.535	0.651
	(0.477)	(0.335)	(0.499)	(0.477)
Drugs	0.412	0.392	0.576	0.815
	(0.492)	(0.488)	(0.494)	(0.388)
Property	0.496	0.491	0.413	0.172
	(0.500)	(0.500)	(0.492)	(0.378)
Violent	0.073	0.099	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	230,361	27,560	152,694	20,626

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

G.7 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** <u>Total</u> the maximum penalties for counts of the primary not scored above

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

[0][ ]

◆ **Additional Offenses** <u>Total</u> the maximum penalties for additional offenses, including counts

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

[0][ ]

◆ **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** <u>Total</u> the maximum penalties for the 5 most recent and serious prior record events

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

[0][ ]

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

◆ **Prior Felony Drug Convictions/Adjudications**

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

[0][ ]

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

◆ **Legally Restrained at Time of Offense**
- None 0
- Other than parole/post-release, supervised probation or CCCA 1
- Parole/post-release, supervised probation or CCCA 4

[0][ ]

*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