How Does Incarceration Affect Reoffending? Estimating the Dose-Response Function

Evan K. Rose

University of Chicago

Yotam Shem-Tov

University of California, Los Angeles

We study the causal effect of incarceration on reoffending using discontinuities in North Carolina's sentencing guidelines. A regression discontinuity analysis shows that 1 year of incarceration causes a reduction in the likelihood of being reincarcerated within 3, 5, and 8 years from sentencing by 44%, 29%, and 21%, respectively. To parse the potentially heterogeneous dose response relationship underlying these effects, we develop an econometric model of prison sentences and recidivism. We find that incarceration has meaningful reoffending-reducing average effects that diminish in incarceration length. As a result, budget-neutral reductions in sentence length combined with increases in incarceration rates can decrease recidivism.

I. Introduction

Since the 1980s, the United States' incarceration rate has more than tripled. The United States now spends $80 billion a year to incarcerate more individuals per capita than any other Organization for Economic

We are particularly indebted to our advisors Patrick Kline, David Card, Steven Raphael, and Christopher Walters for invaluable guidance and support on this project. We thank Magne Mogstad and six anonymous referees for their helpful and constructive comments. We thank Raj Chetty, Avi Feller, Robert Gregory, Hilary Hoynes, Gabriel Lenz, Nicholas Li,

Electronically published October 8, 2021

Journal of Political Economy, volume 129, number 12, December 2021.

https://doi.org/10.1086/716561

Cooperation and Development country. Although crime has steadily declined since the early 1990s, it is unclear to what extent incarceration has contributed to this decrease, since it can impact reoffending through several channels (Kyckelhahn 2011; Lofstrom and Raphael 2016). Prison temporarily incapacitates individuals, removing them from society and making it more difficult to commit crime. In addition, time behind bars can rehabilitate (Kuziemko 2013; Bhuller et al. 2020) and deter (Becker 1968; Drago, Galbiati, and Vertova 2009) offenders or, alternatively, serve as a "school for crime" (Bayer, Hjalmarsson, and Pozen 2009; Stevenson 2017) and break ties to legal labor markets (Grogger 1995; Kling 2006; Raphael 2014; Mueller-Smith 2015; Agan and Starr 2018; Looney and Turner 2018). The balance of effects may depend on the duration of the sentence as well as the individual offender.

This paper studies the causal effect of incarceration on reoffending. We first use a regression discontinuity (RD) design and two decades of administrative data to study the overall effects of harsher sentences. This analysis presents two-stage least squares (2SLS) estimates of the effects of exposure to prison on reoffending and reincarceration in the years after sentencing. We then use a Roy (1951)–style selection model to parse the potentially heterogeneous dose-response function underlying these effects. Using minimal assumptions, we estimate informative bounds on the impact of exposure to different periods of incarceration (e.g., 1 year vs. 3 years) for important populations, such as the average offender. We also bound the impacts of policy-relevant counterfactuals, such as budget-neutral changes in the distribution of sentence lengths.

Our research design isolates exogenous variation in incarceration using discontinuities in North Carolina's sentencing guidelines. These guidelines define permissible punishments according to the convicted offense's severity and a numerical criminal history score. Guideline sentences change discretely at critical score thresholds, shifting sentences

Juliana Londoño-Vélez, Justin McCrary, Conrad Miller, Allison Nichols, Jonathan Roth, Emmanuel Saez, Andres Santos, Jasjeet S. Sekhon, and Danny Yagan for helpful comments and discussions. We thank Bocar Ba and Sam Norris for helpful and constructive comments as conference discussants. We also thank conference and seminar participants at Harvard University; Opportunity Insights; the University of California, San Diego; the University of California, Los Angeles; the University of Michigan at Ann Arbor; the Chicago Crime Lab; the Society of Labor Economics Annual Meeting 2019; the Conference on the Economics of Crime and Justice 2019; the University of California, Irvine; the University of Chicago Economics; the University of Chicago Harris School of Public Policy; the 28th Annual Meeting of the American Law and Economic Association; the University of California, Berkeley Labor Seminar; the University of California, Berkeley Public Finance Lunch Seminar; the All California Labor Conference 2018; and the 13th Annual Conference on Empirical Legal Studies for helpful comments. This paper supersedes a previous draft titled "Does Incarceration Increase Crime?" We gratefully acknowledge financial support from the Center for Equitable Growth. Shem-Tov also acknowledges funding from the US Bureau of Justice Statistics. This paper was edited by Magne Mogstad.

for otherwise comparable individuals. For example, offenders convicted of first-degree burglary face a 30 percentage point jump in the likelihood of incarceration between four and five criminal history points—a difference that can arise because of idiosyncratic factors, such as whether two prior misdemeanors were disposed in the same or consecutive calendar weeks. Although convicted charges are potentially manipulable through plea bargaining, our results are robust to using either the arraigned, the charged, or the convicted offense to define the instruments.

Our RD analysis utilizes multiple important thresholds in North Carolina's guidelines. These discontinuities generate large shifts in prison exposure along both the extensive margin (any prison vs. a probation sentence) and the intensive margin. Combining all our variation using 2SLS, we find that 1 year of incarceration reduces the likelihood of arrests for any new offense by 6.65 percentage points (↓13%), a new violent crime by 2.82 percentage points (↓22%), a new property offense by 2.22 percentage points (↓11%), and a new drug offense by 1.33 percentage points (↓6%) and of being reincarcerated by 14.76 percentage points (↓29%) over the 5 years after sentencing. This reduction in reoffending is still evident even 8 years after sentencing. At this point, offenders sentenced to 1 year of prison are 8.5% and 17% less likely to have ever been arrested for a felony offense or a violent crime, respectively, and are 21% less likely to have ever returned to prison.

To explore the dynamics of these effects and the role of incapacitation, we estimate the impacts of being sentenced to incarceration on offending and incarceration status separately for each month after sentencing. Incarceration sentences naturally generate an immediate spike in the likelihood of being incarcerated that steadily declines over the following months as some individuals are released and others who were not initially incarcerated either reoffend or are imprisoned for violating the conditions of their probation sentence. When effects on incarceration status are more positive, effects on monthly offending rates are correspondingly more negative. Three to eight years after sentencing, those initially incarcerated are no more likely to be incapacitated than those who were not. Monthly offending rates for the two groups are indistinguishable (i.e., there are no effects on "flow" measures of offending). However, incarceration still causes a reduction in cumulative measures of crime, such as ever reoffending in the 8 years after sentencing (i.e., in the "stock" of reoffending).

While informative, the 2SLS estimates do not address several important issues. First, treatment effects are likely to be nonlinear in the duration of exposure, or the "dose." For example, the first 3 months of incarceration may have a very different impact than the last 3 months of a 5-year sentence. Second, treatment effects are likely to be heterogeneous across individuals. Our 2SLS estimates capture weighted average impacts across different doses and different sets of compliers (Angrist and Imbens 1995), essentially aggregating the effects of different treatments on different people. Interpreting just- and overidentified 2SLS estimates is therefore difficult unless treatment effects are in fact linear and homogeneous. Overidentification tests in our basic 2SLS models clearly reject this null. Moreover, using only discontinuities that shift intensive margin exposure to lengthy prison spells produces meaningfully larger reductions in crime per month of prison exposure than using those that shift both the extensive and the intensive margin. Simple fixes that allow for a nonlinear dose response but rule out treatment effect heterogeneity, such as adding polynomial terms for sentence length to the 2SLS model, show the opposite pattern: short sentences generate larger reductions per month.

In the second part of our study, we develop an econometric model that allows us to account for these issues and unpack the nonlinearities and heterogeneity underlying the 2SLS evidence. We model treatment assignment to discrete doses of incarceration as an ordered choice problem that depends on a single unobserved factor (Heckman, Urzua, and Vytlacil 2006). We then extend Mogstad, Santos, and Torgovitsky's (2018) method to the ordered treatment setting and estimate bounds on key parameters (e.g., the average treatment effect [ATE] of a 1-year prison sentence) that are consistent with our reduced-form evidence and plausible restrictions on how unobservables and outcomes are related. These bounds allow for rich dependence of mean outcomes on unobservables and treatment but avoid any distributional assumptions on unobservables or the common assumption of additive separability between observables and unobservables.

The results show that the ATEs of incarceration consist of large reductions in reoffending. The average ATE for offenders with observables that place them at the five most important sentencing guideline discontinuities indicate that 3 years of prison reduce the likelihood of reincarceration within 5 years of sentencing by 38–56 percentage points. Reductions from the first year of exposure are roughly twice as large as from the second or third year. Bounds that examine effects on the populations at each discontinuity separately show similar results. We also find evidence of selection into treatment and heterogeneous treatment effects. Offenders most likely to be given a harsh prison sentence are also the most likely to reoffend. However, treatment effects for the most hardened criminals are also the largest, indicating that judges most harshly punish offenders for whom incarceration causes the largest reductions in reoffending.

Motivated by our findings, we conclude by using the selection model to examine the impact of budget-neutral counterfactual changes in sentencing policy. These counterfactuals reduce average sentences and use the savings to give more offenders short (<1-year) prison spells. Since we find that incarceration's impacts on ever reoffending are largest for the initial exposure, such reallocations might reduce average reoffending rates. However, since we also find that treatment effects reduce reoffending the most for those currently sentenced to the longest spells, the full impact is ambiguous. The results indicate that such reallocations are beneficial in the best case and not damaging in the worst case. For offenders convicted of moderately severe crimes (e.g., robbery or theft), for example, we find that reducing average sentence lengths by roughly 50% and incarcerating nearly all offenders for at least 3 months could reduce reincarceration rates by nearly 3.5 percentage points and the cumulative number of days reincarcerated by more than 30 days per offender.

We contribute to a broad literature across the social sciences on the relationship between incarceration and reoffending.[1] In recent years, a common empirical strategy has been to take advantage of random or rotational assignment of defendants to judges.[2] A few recent papers utilizing this design have found results broadly consistent with this analysis. Bhuller et al. (2020), for example, find that prison sentences have substantial rehabilitative effects among Norwegian criminal defendants. In the United States, Norris, Pecenco, and Weaver (2020) find that incarceration sentences cause a long-run reduction in reoffending using data from three large counties in Ohio. Our estimates are similar in sign but smaller in magnitude than Bhuller et al. (2020) and are broadly comparable in both sign and magnitude to Norris, Pecenco, and Weave (2020). They differ, however, from Mueller-Smith (2015), who finds that exposure to prison increases reoffending using data from Harris County, Texas. In addition to providing new evidence, we build on and extend Mueller-Smith (2015), Bhuller et al. (2020), and Norris, Pecenco, and Weave (2020) in several ways. The multiple discontinuities we exploit provide variation in both the extensive and the intensive margin effects of incarceration, allowing us to estimate nonlinearity in the effects of incarceration on reoffending. In addition, our selection model allows us to bound effects for the average offender and for other policy-relevant populations, rather than just the compliers for our instruments, while correctly accounting for nonlinearity and unobserved

[1] The majority of the previous literature focused on the incapacitation channel. Notable examples include Levitt (1996), Owens (2009), Buonanno and Raphael (2013), Barbarino and Mastrobuoni (2014), and Raphael and Lofstrom (2016). Miles and Ludwig (2007) provide a review of the evidence from the criminology literature.

[2] Examples of papers using a judges design to obtain exogenous variation in sentences and intermediate case outcomes (e.g., bail) include Kling (2006), Green and Winik (2010), Loeffer (2013), Nagin and Snodgrass (2013), Aizer and Doyle (2015), Mueller-Smith (2015), Stevenson (2016), Harding et al. (2017), Arnold, Dobbie, and Yang (2018), Bhuller et al. (2018, 2020), Dobbie, Goldin, and Yang (2018), Dobbie et al. (2018), Norris (2018), Huttunen et al. (2019), Aneja and Avenancio-Leon (2020), Arteaga (2020), Norris, Pecenco, and Weave (2020), and Zapryanova (2020).

heterogeneity. These factors may be important drivers of differences across studies in the literature.[3]

Several papers exploiting non–judge variation also find results similar to those in this paper. Kuziemko (2013), for example, compares a parole system with a fixed-sentence regime and argues that each additional month in prison reduces 3-year reincarceration rates by 1.3 percentage points for a sample of parolees in Georgia. On the other hand, Franco et al. (2020) find that reincarceration rates are higher for initially incarcerated offenders.[4] Differences in the institutional setting and the impact of accounting for violations of technical rules imposed on probation and parole can potentially explain some of these differences. We discuss this issue in section II.D and propose possible solutions. A final strand of related literature uses exogenous shocks to prison populations to identify the relationship between incarceration rates and crime.[5] This type of variation captures effects that go beyond the partial equilibrium analysis we study in this paper. Estimates from this literature vary but generally find that decreases in incarceration rates generate increases in at least some categories of crime (Levitt 1996; Raphael and Lofstrom 2016).

The remainder of this paper is organized as follows. Section II describes the institutional setting and the data used. Section III describes the empirical strategy for identifying causal effects and reports results from the 2SLS analysis. Section IV lays out the selection model and our strategy for estimating bounds on relevant parameters of interest and reports the results of this approach. Section V discusses some of the policy implications of our results by estimating the impacts on reoffending of budget-neutral counterfactual sentencing policies. Section VI concludes.

II. Setting and Data

In this section, we describe the sentencing guidelines that determine felony punishments in North Carolina and are the source of our instrumental variation. We also describe the sources of our data, detail how we construct our primary analysis sample, and provide summary statistics.

[3] For example, Estelle and Phillips (2018) find that harsher sentences reduce drunk drivers' reoffending when using variation from sentencing guidelines but not when using variation from judge assignment.

[4] Studies on juvenile offenders also find mixed results (Levitt 1998; Hjalmarsson 2009; Aizer and Doyle 2015). However, the effects of incarceration may be different for juvenile vs. adult felony offenders, who are our focus.

[5] Notable examples include Marvell and Moody (1994), Levitt (1996), Kessler and Levitt (1999), Drago, Galbiati, and Vertova (2009), Maurin and Ouss (2009), McCrary and Sanga (2012), Buonanno and Raphael (2013), Barbarino and Mastrobuoni (2014), and Raphael and Lofstrom (2016).

A. *Structured Sentencing in North Carolina*

Our research design relies on the structure of North Carolina's mandatory sentencing guidelines, which were first introduced on October 1, 1994, by North Carolina's Structured Sentencing Act (hereafter, SSA). These guidelines were crafted as part of a nationwide shift toward rule-based criminal sentencing motivated by a desire to reduce sentencing disparities across judges and defendants and to limit discretion in the sentencing and parole process. In 1996, 16 states had sentencing guidelines and 20 had some form of deterministic sentencing (US Department of Justice 1996). By 2008, the number of states with sentencing guidelines had increased to 28 (National Center for State Courts 2008).[6]

The SSA eliminated parole by requiring that defendants serve the entirety of a minimum sentence.[7] The law established separate misdemeanor and felony "grids" that determine these minimum sentences as a function of offense severity and the offender's criminal history.[8] Felony offenses are grouped into 10 different classes based on the severity of the offense. Offenders are assigned a criminal history score (referred to as "prior record points") that assigns one point for misdemeanor offenses and two to 10 points for felony offenses, depending on the seriousness of the crime. When an individual was previously convicted of multiple offenses in the same calendar week, only the most serious offense is used.[9] Additional points are added if offenses are committed while the offender is on probation or if the current offense is sufficiently similar to any prior offenses. As a result of these details, two individuals with highly similar criminal histories can have different prior record scores depending on the timing and precise nature of their previous offenses.[10]

The SSA groups individuals into prior record "levels" according to their total prior points and sets minimum sentences for each offense class and

[6] Sentencing guidelines have been used elsewhere to estimate effects of features of the criminal justice system. Ganong (2012) and Kuziemko (2013) study the case of parole, Hjalmarsson (2009) studies juvenile offenders, and Chen and Shapiro (2007) study prison conditions. In Michigan, Estelle and Phillips (2018) and Harding et al. (2018) use similar designs to examine the effects of different criminal sanctions (e.g., prison vs. probation) on recidivism.

[7] After doing so, defendants become eligible for early release but can serve no more than 120% of their minimum sentence. Figure A.1 (figs. A.1–A.10, C.1–C.7, D.1, D.2, and E.1–E.5 are available online) shows the relationship between the minimum sentenced incarceration length and the actual number of months served incarcerated.

[8] Driving while impaired (DWI) and drug trafficking offenses have separate sentencing guidelines.

[9] Of the offenders in our analysis data set, 10.8% were convicted of multiple different offenses within 5–10 days.

[10] For these reasons, individuals may not know their exact number of prior points, which are officially calculated only at sentencing. We do not find any evidence of discontinuities in the density of offenders above or below critical prior point thresholds that determine sentencing severity (fig. C.6).

	I 0 Pts	II 1-4 Pts	III 5-8 Pts	IV 9-14 Pts	V 15-18 Pts	VI 19+ Pts
E	**I/A**	**I/A**	**A**	**A**	**A**	**A**
	25 - 31	*29 - 36*	*34 - 42*	*46 - 58*	*53 - 66*	*59 - 74*
	20 - 25	**23 - 29**	**27 - 34**	**37 - 46**	**42 - 53**	**47 - 59**
	15 - 20	*17 - 23*	*20 - 27*	*28 - 37*	*32 - 42*	*35 - 47*
F	**I/A**	**I/A**	**I/A**	**A**	**A**	**A**
	16 - 20	*19 - 24*	*21 - 26*	*25 - 31*	*34 - 42*	*39 - 49*
	13 - 16	**15 - 19**	**17 - 21**	**20 - 25**	**27 - 34**	**31 - 39**
	10 - 13	*11 - 15*	*13 - 17*	*15 - 20*	*20 - 27*	*23 - 31*
G	**I/A**	**I/A**	**I/A**	**I/A**	**A**	**A**
	13 - 16	*15 - 19*	*16 - 20*	*20 - 25*	*21 - 26*	*29 - 36*
	10 - 13	**12 - 15**	**13 - 16**	**16 - 20**	**17 - 21**	**23 - 29**
	8 - 10	*9 - 12*	*10 - 13*	*12 - 16*	*13 - 17*	*17 - 23*
H	**C/I/A**	**I/A**	**I/A**	**I/A**	**I/A**	**A**
	6 - 8	*8 - 10*	*10 - 12*	*11 - 14*	*15 - 19*	*20 - 25*
	5 - 6	**6 - 8**	**8 - 10**	**9 - 11**	**12 - 15**	**16 - 20**
	4 - 5	*4 - 6*	*6 - 8*	*7 - 9*	*9 - 12*	*12 - 16*
I	**C**	**C/I**	**I**	**I/A**	**I/A**	**I/A**
	6 - 8	*6 - 8*	*6 - 8*	*8 - 10*	*9 - 11*	*10 - 12*
	4 - 6	**4 - 6**	**5 - 6**	**6 - 8**	**7 - 9**	**8 - 10**
	3 - 4	*3 - 4*	*4 - 5*	*4 - 6*	*5 - 7*	*6 - 8*

DISPOSITION
Aggravated Range
PRESUMPTIVE RANGE
Mitigated Range

Fig. 1.—This figure shows the sentencing guidelines, or “grid,” applicable to offenses committed after December 1, 1995, but before December 1, 2009. Appendix B includes the full set of guidelines from 1995 to the present. Each offense is classified to a severity class that determines the applicable row of the grid. Offenders receive a numerical criminal history score, or “prior points,” which is a weighted sum of prior convictions based on severity and timing, that determines the applicable column. The columns group multiple prior point values into a prior record level. The numbers in each offense class and prior record level “cell” define minimum incarceration sentences. Maximum sentences are always 120% of the minimum. Sentences are specified for three different ranges: aggravated, presumptive, and mitigated. Each cell is assigned a set of recommended sentence types: “A” denotes active incarceration, and “C” and “I” denote the type of probation. When a nonincarceration sentence is imposed, the incarceration sentence recommended by the grid is suspended. Probation sentences are typically between 18 and 36 months. The thick black lines indicate places in the grid where recommended sentence types change. Indicators for having offense class and prior point combinations that fall to the right of each thick black line comprise our primary instruments. A color version of this figure is available online.

prior record level combination, which we refer to as a grid “cell.”[11] This is illustrated in figure 1, which shows the portion of North Carolina’s sentencing grid that we study. Each grid cell has a set of allowable sentence types: (i) active punishment (state prison or jail); (ii) intermediate punishment, which is probation with at least one of several possible special conditions;[12] and (iii) community punishment, or regular probation. These

[11] The maximum and minimum sentences are specified for three different ranges: aggravated, presumptive, and mitigated. The majority of crimes are sentenced in the presumptive range. The determination of sentencing range is independent of criminal history. For example, if a defendant convicted of a class E offense is in the aggregated range when his prior record level is III, then she would also be sentenced in the aggravated range if she had a prior record level of II or IV.

[12] Intermediate can also include “shock” probation, which includes a short incarceration spell before probation begins.

sentence types are denoted with "C/I/A" lettering at the top of each cell in the grid.[13] The thick black lines demarcate thresholds in the grid where the set of allowable punishment types changes to either include prison time or exclude nonprison sentences. The numbers in the grid specify ranges for potential incarceration sentences only. When an offender receives a nonincarceration punishment, this sentence is suspended. Probation sentences are required to be between 18 and 36 months except under special circumstances.

The combination of shifts in required sentence lengths and allowable sentence types generates large differences in recommended punishments across the grid, as shown in figure 1. For example, offenders with nine prior points and a class I conviction can be given an incarceration sentence, whereas offenders with eight points cannot. Because individuals are usually sentenced at the bottom of the grid ranges, moving between cells generates meaningful changes in the intensive margin as well. The grid has been modified occasionally since its introduction, which also generates variation in sentences. We exploit one such reform in 2009 that substantially modified the mapping between prior record points and grid cells to validate our research design.

B. Data Sources

We use administrative information on arrests, charges, and sentencing from two sources. The first consists of records provided by the North Carolina Administrative Office of the Courts (AOC) covering 1990–2017. These data include rich information on defendants, offenses, initial charges, convictions, and sentences for all cases disposed in North Carolina Superior Court, which hears felony cases. These data are used to measure the set of initial charges associated with a conviction and to construct some reoffending measures. Because criminal charges in North Carolina are initially filed by law enforcement officers (as opposed to prosecutors), the charges in these data closely approximate arrests for offenses that would be heard in superior court. We date new charges (or convictions) using the date of offense, rather than the date that charges were filed, to eliminate any delays owing to lags in detection or in our court proceedings.

Second, we use records from the North Carolina Department of Public Safety (DPS) that contain detailed information on the universe of individuals who received supervised probation or incarceration sentences from the 1970s to the present. These data allow us to observe sentencing inputs and outcomes, including the severity class of each felony offense, prior record points, and ultimate sentences. The data also contain data

[13] For more details, see the official sentencing guidelines for the years 1994–2013 in app. B (apps. A–F are available online).

on probation revocations and additional details on offenders' demographics, including age, height, weight, languages spoken, race, and ethnicity. We use these data to construct our instruments and to measure treatment. When studying new arrests as an outcome, we take the union of incidents recorded in either data set to provide the most complete coverage of criminal activity possible.

C. *Sample Construction and Restrictions*

Because our research design utilizes discontinuities in felony sentencing guidelines, the analysis sample is restricted to individuals convicted of felony offenses committed between 1995 and 2014 and therefore sentenced on the felony grid. We do not include misdemeanors or DWIs, since they are sentenced under different guidelines. We focus on class E through class I offenses (92.3% of the observations) and include individuals with fewer than 25 prior record points. While classes more severe than E (e.g., classes D and C) also have discontinuities that affect incarceration on the intensive margin, they do not have discontinuities that affect whether prison sentences are allowed at all.[14] Finally, we also restrict the analysis to individuals aged between 18 and 65 at the time of offense.

Offenders routinely face multiple charges simultaneously and can be sentenced to concurrent incarceration spells for offenses committed at different dates. To overcome this issue, we conduct our analysis at the charge/offense level and cluster standard errors by individual. When an offender has several charges that were sentenced jointly and thus have corresponding incarceration spells that begin at the same time, we keep only the most severe charge, since the sentences are concurrent and the most severe charge determines the spell length.[15]

D. *Measuring Reoffending*

Our primary reoffending measure is an indicator for whether an individual is incarcerated within a fixed time horizon from the date of sentencing.

[14] Including classes C and D in the analysis does not alter any of our results. Table A.1 (tables A.1–A.16, E.1, and E.2 are available online) lists the five most frequent offenses of individuals in our sample by their convicted severity class. For example, offenders in class I (least severe offense class) are most frequently convicted of possession, forgery, and breaking and entering vehicles. Offenders in class E (most severe offense class) are most commonly convicted of assault with a deadly weapon and second-degree kidnapping.

[15] Another approach would be to group charges into cases where either the conviction, the offense, or the sentencing dates of offenses fall within a certain time period (e.g., 30 days) from each other. We have experimented with a variety of different grouping methodologies; the results from all strategies are similar. The main difference is how accurately each grouping method estimates the actual time served for a given offense. We found that the charge-level approach we use most accurately measures the length of time the individual served in prison for each offense.

Reincarceration is commonly used as a measure of recidivism (e.g., Kuziemko 2013; Yang 2017; Agan and Makowsky 2021). In addition, we also consider indicators for being arrested for any new offense, new offenses of different types (e.g., violent vs. property crimes), and counts of total new offenses or days spent in prison for new offenses.

Offenders not sentenced to incarceration are instead sentenced to probation. While on probation, offenders face restrictions on alcohol and drug use, work and socializing, and travel and are required to pay off court fees and fines. Violating these restrictions can lead to probation being revoked and incarceration. An important secondary decision, therefore, is whether to count probation violations as new offenses and whether to include any resulting incarceration spells.[16] Our measure of reincarceration includes being incarcerated for both new offenses and probation violations. To ensure that our results are not overly sensitive to this decision, when studying new arrests as an outcome we do not count probation violations. In our robustness checks and appendix material, we provide a variety of other tests for the robustness of our results to how we handle this issue. This includes estimates that assume probation revocations cause censoring at random—in other words, that the risks of probation revocations and new offenses are independent. Under this assumption, we can simply drop any observations for which a probation revocation occurred before a new offense. In practice, we view these independent risk estimates as an upper bound, since it seems unlikely that probation revocations are negatively correlated with risk—that is, that the least dangerous individuals are most likely to be revoked for technical violations of probation.[17]

E. Summary Statistics

Summary statistics for our sample are presented in table 1. On average, offenders are predominately male, roughly 50% black, and 31 years old (median: 28) at the time they committed their offense. More than two-thirds of cases do not result in prison or jail sentences. Incarceration sentences average about 4.7 months. Conditional on receiving an incarceration sentence, the average length is 13.2 months. Roughly 57% of the sample reoffends at some point in the period we study. Most offenders

[16] These technical revocations are frequently not associated with an arrest for a new criminal offense. However, probation officers may also revoke individuals they suspect are involved in new criminal activity. For example, Austin and Lawson (1998) find that in California most technical violations of parole were associated with a new criminal offense that was not prosecuted. This scenario is frequently mentioned as a motivation for counting probation revocations as reoffending, although many studies do not discuss the issue explicitly.

[17] Estimates from Rose (2021) show that the risk of revocation is positively correlated with the risk of reoffending, implying that the "true" effects on reoffending in a regime without probation may lie closer to estimates including probation revocations in the reoffending measure.

TABLE 1
SUMMARY STATISTICS: DEMOGRAPHICS, SENTENCING, AND REOFFENDING

	FULL ANALYSIS SAMPLE		RD WINDOW ONLY	
	Mean (1)	Median (2)	Mean (3)	Median (4)
Demographics:				
Male	.81		.88	
Race:				
White	.43		.38	
Black	.50		.58	
Other	.07		.038	
Born in North Carolina	.69		.74	
Age at offense	30.63	28.00	33.46	32.00
Age at conviction	31.62	29.46	34.47	33.15
Incarceration measures:				
Sentenced to any incarceration	.35		.50	
Incarceration sentence (months)	4.72	.00	7.80	.33
Months served (months)	6.29	.00	10.70	.03
Incarceration sentence conditional on positive sentence (months)	13.15	10.00	15.36	15.00
Months served conditional on positive sentence (months)	20.41	11.44	21.38	15.29
Recidivism measures from sentencing:				
Recidivate in 1 year	.16		.15	
Felony recidivate in 1 year	.10		.18	
Recidivate in 2 years	.28		.29	
Felony recidivate in 2 years	.18		.18	
Recidivate in 3 years	.37		.39	
Felony recidivate in 3 years	.25		.27	
Recidivate in 5 years	.47		.51	
Felony recidivate in 5 years	.33		.37	
Recidivate in period	.57		.62	
Felony recidivate in period	.44		.48	
Days to recidivate from conviction conditional on recidivating	1,073.76	741.00	1,088.86	805.00
Total observations	517,091		102,839	
Total unique individuals	314,538		78,983	

NOTE.—This table shows summary statistics for the primary analysis sample and the sample close to the discontinuities we use in the majority of our analysis. Not all observations are included in all regressions, since when using outcomes measured over a fixed horizon (e.g., reoffending within 3 years of sentencing) we restrict the sample to observations observed over that horizon. This drops some observations sentenced toward the end of the sample period. The difference between average sentences and average months served reflects both the fact that sentences represent minimum sentences and the fact that offenders may face multiple consecutive or concurrent sentences. The unit of analysis in our sample is an individual–sentencing date pair. When an offender has several charges that were sentenced jointly and thus has corresponding incarceration spells that begin at the same time, we keep only the most severe charge, since the sentences are concurrent and the most severe charge determines the spell length. Columns 1 and 2 describe the full analysis sample, and cols. 3 and 4 describe the observations in the RD window—i.e., that are located in grid cells adjacent to a punishment type discontinuity.

who reoffend do so in the first few years after being convicted. Forty-seven percent of offenders reoffend within 5 years of sentencing, and 28% reoffend in the first 2 years. Overall, the sample is typical for offenders at risk of an incarceration sentence; North Carolina has incarceration and recidivism rates similar to the US average.[18]

Columns 3 and 4 of table 1 report summary statistics for individuals in grid cells adjacent to one of the five punishment type discontinuities, where the set of allowable sentences changes and which comprise the primary instruments used in our analysis. For example, this sample includes individuals with prior record levels I and II in class E. These observations directly contribute to our estimated effects of incarceration when using these discontinuities as instruments. Their characteristics are thus the most relevant for our estimates. Offenders in this sample are slightly older (median age: 32), less likely to be white (38% relative to 43%), and more likely to be born in North Carolina (75% relative to 69%).

III. Causal Effects of Incarceration

In this section, we estimate the effect of incarceration on reoffending. We begin by describing our empirical strategy in section III.A, present reduced-form estimates in section III.B, and discuss 2SLS estimates of the effects of incarceration length in section III.C. In section III.D, we investigate nonlinearity and unobserved heterogeneity in the effects of incarceration. Finally, in section III.E we show a variety of robustness checks that reinforce the causal interpretation of our estimates.

A. Empirical Strategy

Our research design exploits nonlinearities in sentencing outcomes at the boundaries of horizontally adjacent sentencing grid cells. With five offense classes (i.e., rows) and six prior record levels (i.e., columns), there are a total of 25 such cell discontinuities. Each SSA cell contains four or five values of the running variable (prior points) except in the first column, which contains just one or two, depending on the year. Our setting is thus not a classic RD scenario with a continuous running variable, such as a congressional election (Lee 2008) or a college loan program (Solis 2017). Instead, we have a discrete running variable; our specification therefore reflects a parameterized RD design (Clark and Del Bono 2016).[19]

[18] See fig. 1 in Norris, Pecenco, and Weave (2020).

[19] Clark and Del Bono (2016) study school district allocation and use nonlinearities in the assignment formula to construct a "parameterized regression kink design." Other studies that utilize nonlinearities in assignment mechanisms include Kuziemko (2013) for the case of parole.

Our model includes separate linear slopes in each SSA cell and allows for vertical shifts—or "jumps"—between horizontally adjacent cells. Figure 2*A* visually illustrates this idea for class F offenses when the outcome is any incarceration. The spaces between each line reflect cell boundaries and thus potential instruments. The large jump at the dotted vertical line reflects the punishment type discontinuity for class F, after which probation is no longer a permissible sentence.

Our preferred estimator uses only the five discontinuities where allowable punishment types change as excluded instruments. The empirical specification stacks the variation from each discontinuity—one in each felony class—to estimate a single treatment effect and is written formally in the two-equation system below. The first stage, equation (1), estimates incarceration length as a function of prior points, convicted charge severity, punishment discontinuities, and other covariates; equation (2) models the relationship between an outcome measured within t months of sentencing, incarceration length, and nonexcluded controls:

$$\begin{aligned} D_i = {} & \underbrace{\eta^1_{\text{class}_i} + X_i'\alpha_1}_{\text{baseline controls}} + \underbrace{\sum_k 1\{\text{class}_i = k\}\left[\sum_l \beta^1_{lk} 1\{p_i \geq l\}(p_i - l + 0.5) + \psi^1_k p_i\right]}_{\text{linear slopes in prior points by class and level}} \\ & + \underbrace{\sum_{k,l \in \text{punish}} \xi_{kl} 1\{p_i \geq l\} 1\{\text{class}_i = k\}}_{\text{punishment type discontinuities}} + \underbrace{\sum_{k,l \notin \text{punish}} \gamma^1_k 1\{p_i \geq l\} 1\{\text{class}_i = k\}}_{\text{other discontinuities}} + \varepsilon_i, \end{aligned} \tag{1}$$

$$\begin{aligned} Y_{it} = {} & \beta_0 D_i + \underbrace{\eta^2_{\text{class}_i} + X_i'\alpha_2}_{\text{baseline controls}} + \underbrace{\sum_k 1\{\text{class}_i = k\}\left[\sum_l \beta^2_{lk} 1\{p_i \geq l\}(p_i - l + 0.5) + \psi^2_k p_i\right]}_{\text{linear slopes in prior points by class and level}} \\ & + \underbrace{\sum_{k,l \notin \text{punish}} \gamma^2_k 1\{p_i \geq l\} 1\{\text{class}_i = k\}}_{\text{other discontinuities}} + e_{it}, \end{aligned} \tag{2}$$

where D_i represents the length of incarceration offender i served, $\eta^1_{\text{class}_i}$ and $\eta^2_{\text{class}_i}$ represent row (i.e., offense class)–specific intercepts, and p_i represents prior points. The thresholds l refer to the prior record boundary levels in place at the time of the offense (e.g., five or nine points). When estimating the changes in slope on either side of each boundary (the $1\{p_i \geq l\}(p_i - l + 0.5)$ terms), we recenter by $l - 0.5$ so that we measure the discontinuity halfway between the boundary prior point values as implied by the linear fits on either side, rather than at either extreme.[20]

[20] This appears to be the most natural choice given the discrete nature of the data, although our results are not sensitive to this decision.

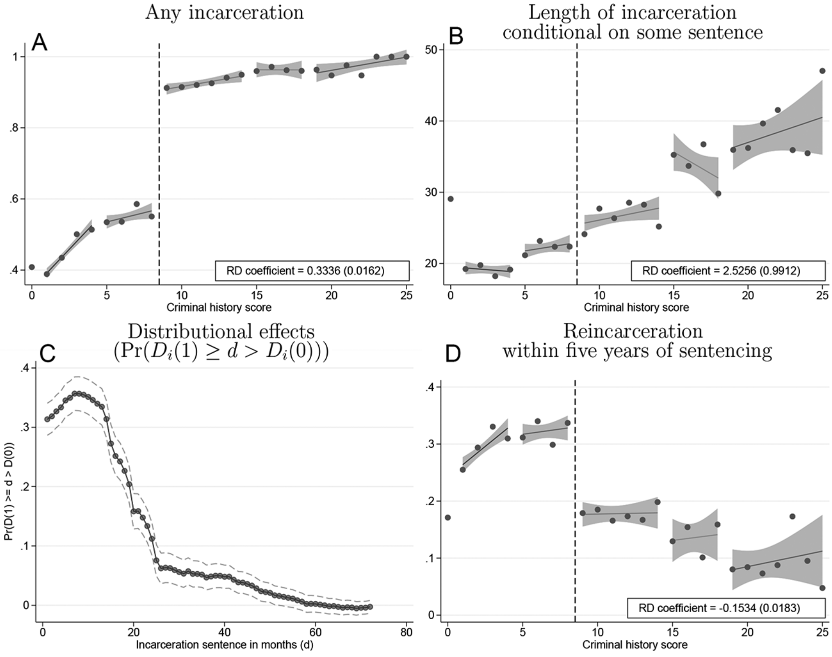

Fig. 2.—This figure shows the first-stage effect of the punishment type discontinuity in class F on any incarceration and on the length of incarceration. In addition, it also demonstrates the reduced-form effect on reincarceration within 5 years of sentencing. *A*, Share of offenders sentenced to any term of incarceration plotted against the running variable, prior record points. *B*, Average sentence of offenders who have been sentenced to some term in prison plotted against the running variable, prior record points. *C*, Estimates of the shifts in incarceration exposure generated by the instrument ($\Pr(D_i(1) \geq d > D_i(0))$), which correspond to the unnormalized weights in the average causal response (Angrist and Imbens 1995). These shifts reflect the probability that an offender would spend fewer than d months incarcerated if assigned $Z_i = 0$ (just below the discontinuity) but at least d months if assigned $Z_i = 1$ (just above the discontinuity). This probability can be estimated nonparametrically using $\mathbb{E}[1(D_i \geq d)|Z_i = 1] - \mathbb{E}[1(D_i \geq d)|Z_i = 0]$, which corresponds to the ξ coefficients in our first-stage specification when $1(D_i \geq d)$ is the outcome. *D*, Reincarceration rate plotted against the running variable. This illustrates the reduced-form impacts of the discontinuities on the likelihood of being reincarcerated. Panels *A*, *B*, and *D* include data only for offenses sentenced under the sentencing grid that applied to offenses committed between 1995 and 2009. In 2009, the guidelines changed and the discontinuities shifted by one prior point either to the left or to the right. All of the official grids are included in appendix B. Similar figures for other classes are shown in figure A.2. Standard errors are clustered at the individual level. A color version of this figure is available online.

The covariates X_i include demographic controls (e.g., age and gender), our own measures of criminal history (e.g., fixed effects for prior convictions), and other controls discussed further below.

While the main analysis uses the five punishment type discontinuities as instruments, we also explore estimates excluding other discontinuities. Since these discontinuities tend to shift offenders along the intensive margin of exposure to prison, they are useful for diagnosing nonlinearities in effects of incarceration, as we discuss below. However, we never use the boundary between the first and second prior record level to generate instruments, since we cannot estimate a slope in prior points on one side.

1. First-Stage Effects of Discontinuities

This research design captures large discontinuities in sanctions across the sentencing grid. For example, figure 2*A* shows that for an offender convicted of a class F felony offense (e.g., assault with serious injury), the probability of incarceration increases by 33 percentage points between eight and nine prior points but varies smoothly elsewhere. Figure 2*B* shows that each discontinuity also generates shifts in sentence lengths (conditional on positive) at every cell boundary. Figure A.2 documents multiple discontinuities in both the type and the length of punishment for all other offense classes. This variation occurs at different values of prior record points depending on the class. For example, in class H, which contains the most defendants in the data, the largest extensive margin shift occurs between prior record points 18 and 19.

To examine how the instruments impact the entire distribution of incarceration length, let D_i denote months of incarceration and Z_i denote whether an individual is above or below a punishment type discontinuity. We estimate $\Pr(D_i(1) \geq d > D_i(0))$ for every level of d and for each of the five punishment type discontinuities. These probabilities are directly estimated by the ξ_{kl} coefficients in equation (1) when the outcome is $1(D_i \geq d)$. Figure 2*C* plots the $\widehat{\Pr}(D_i(1) \geq d > D_i(0))$ estimates for class F, with the remaining offense classes in figure A.3. The instruments provide substantial variation in exposure to incarceration. Each offense class also provides quite different variation, with some classes concentrated on short durations and others generating shifts in durations beyond 2 or 3 years.[21]

In the regressions that follow, we control for offenders' criminal history using both the linear controls in prior points from the RD specification and indicators for the number of previous incarceration spells, the number of previous convictions, and fixed effects for the months spent incarcerated prior to the current conviction. Even after taking into account criminal history, the grid boundaries still provide strong variation in the type and length of punishment, as shown by the first-stage F statistics presented in the tables of results that follow. The instrumental variation

[21] As noted by Angrist and Imbens (1995), estimates of $\Pr(D_i(1) \geq d > D_i(0))$ also provide a test for the monotonicity assumption. If the instruments satisfy monotonicity, then $\Pr(D_i(1) \geq d > D_i(0))$ should never cross the X-axis at zero, since a probability cannot have a negative value. Figure A.3 confirms that all the instruments pass this validity check.

therefore primarily comes from the nonlinear mapping between prior convictions and prior record points, as opposed to simple counts of prior convictions.

2. Instrument Validity

We perform a series of balance and validation exercises to assess the validity of the instruments. These analyses demonstrate that our instruments do not predict individual characteristics, supporting the assumption that changes in outcomes at each discontinuity reflect the effects of incarceration rather than selection. Since there are many relevant pretreatment covariates, we make use of a predicted reoffending score calculated by regressing an indicator for reoffending on all the pretreatment covariates among only nonincarcerated offenders and fitting predicted values.[22]

Figure 3 shows that the predicted risk score evolves smoothly across each of the five punishment type discontinuities. In each case, the changes at the discontinuity are negligible. A test for the joint significance of all five discontinuities also fails to reject zero effects (the P-value is .159, with an F statistic of 1.58 and five degrees of freedom). The smoothness of offenders' covariates across thresholds is especially encouraging in light of the large first-stage discontinuities in sentences documented in figure 2. Figures C.1 and C.2 show that specific covariates, such as the offender's age at the time of offense and previous incarceration history, also evolve smoothly across the discontinuities in the sentencing guidelines.

Several additional analyses further support the validity of our design. First, for every covariate, we measure the difference in means between each pair of consecutive prior points within a grid row. The overall distribution of these differences is not distinguishable from the difference in means between the points straddling the discontinuities (see fig. C.3). In other words, although sentences change abruptly across consecutive prior points at the discontinuities in punishment type (see fig. C.4), other observable characteristics do not.

Second, we use a 2009 sentencing reform that shifted each discontinuity one prior point to the left or right, depending on the offense class. This change shifted punishments as well, as shown in figure A.4. Despite this shift, the distribution of offenders' covariates across prior points remained the same, indicating no scope for sorting across discontinuities.

[22] Summarizing imbalance by the covariates' relationship to the outcome surface is a common methodology in the literature (e.g., Bowers and Hansen 2009; Londoño-Vélez, Rodríguez, and Sánchez 2020). We also experimented with using more sophisticated (i.e., machine learning models) to construct the risk score; the results are similar.

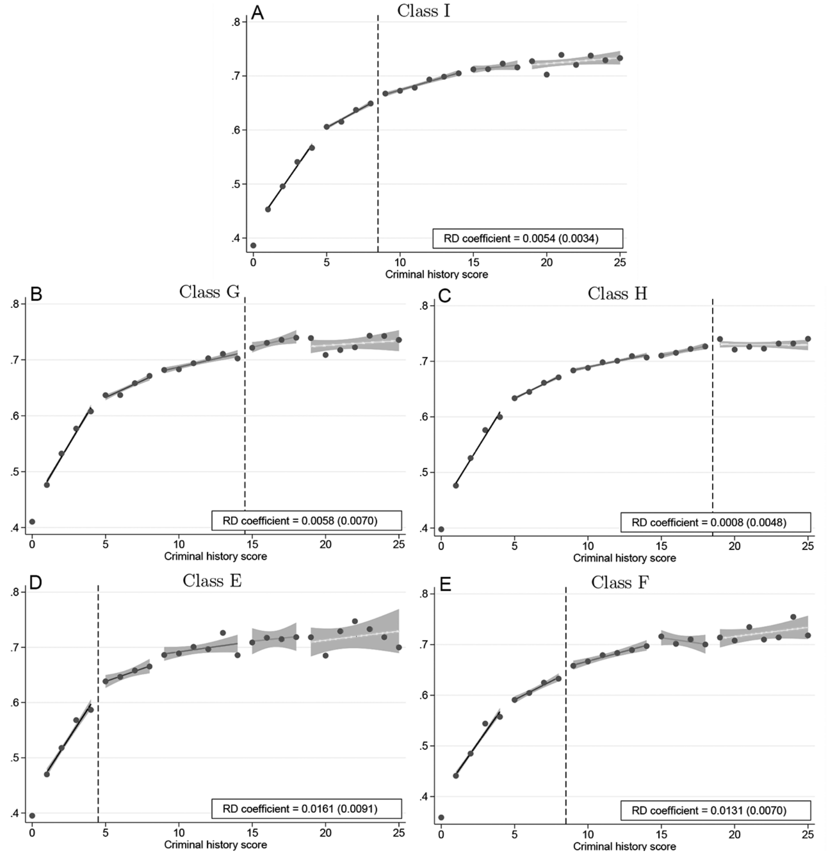

FIG. 3.—This figure demonstrates that a summary index of the covariates varies smoothly across sentencing grid discontinuities. We calculate the predicted values from a simple linear regression of all available covariates (e.g., age, race, criminal history) on reoffending within 5 years of the time of release (using only nonincarcerated offenders). Each panel plots the mean of this index against prior points for each offense class separately. The dotted lines reflect the punishment type discontinuities that comprise our primary instruments. We use a summary index because there are many potentially important pretreatment covariates. Summarizing imbalance by the covariates' relationship to the outcome surface is a common methodology in the literature (e.g., Bowers and Hansen 2009; Londoño-Vélez, Rodríguez, and Sánchez 2020). Standard errors are clustered at the individual level. Only offenses sentenced under the sentencing grid that applied to offenses committed between 1995 and 2009 are plotted. A color version of this figure is available online.

We demonstrate this by estimating equation (1) in the 2 years before and after the change but define the location of each discontinuity using the prereform grid. We then interact indicators for being to the right of each discontinuity with an indicator for being sentenced under the new grid and test for their joint significance. As shown in table A.2, these interactions strongly predict changes in incarceration exposure, but we cannot reject the null that that risk scores and individual covariates are unchanged after the reform. Large changes in punishments therefore do not lead to changes in sorting along observable dimensions.[23]

Finally, figure C.6 shows that there is no evidence of discontinuities in the density of offenders at the punishment type discontinuities. Figure C.7 reports the results of a McCrary (2008)–style test and shows that the changes in the density at the discontinuities are not distinguishable from zero and are not correlated with changes in the likelihood of incarceration. Overall, therefore, there is strong support for the validity of our instruments. Nevertheless, after estimating our core results, we conduct additional robustness checks to further support this claim and investigate other potential concerns, such as sorting through plea bargaining and differences in the likelihood of criminal activity being detected while on probation.

B. Reduced-Form Estimates

We begin by studying the reduced-form effects of our instruments on reoffending outcomes. Figure 2*D* illustrates these effects for class F when the outcome is an indicator for being reincarcerated within 3 years of sentencing. At the punishment type discontinuity between nine and 10 prior points, for example, reincarceration rates fall by 15.3 percentage points. Reincarceration rates also fall at other discontinuities, where the shifts in incarceration exposure primarily fall on the intensive margin. To summarize the evidence from all of our instruments, we estimate reduced-form effects using equation (1) but imposing that the coefficients on indicators for being to the right of a punishment type discontinuity are all equal (i.e., $\xi_{E,4} = \xi_{F,9} = \xi_{G,14} = \xi_{H,19} = \xi_{I,9} = \xi^{RF}$). This strategy averages effects across all five offense classes, collapsing our variation into a single coefficient.[24]

We first consider indicators for being incarcerated or for reoffending within a given month over the 8 years after sentencing. These estimates are plotted in figure 4, where each point in panel *A* represents a separate estimate of ξ^{RF} for each outcome measured at the point in time on the *X*-axis. Figure 4 shows that the discontinuities cause a large and immediate increase in incarceration, which confirms the strength of our first stage. The effect declines steadily over the following months as some individuals are released and others who were not initially incarcerated either reoffend or have their probation revoked. After approximately 2.5 years, the effect

[23] Figure C.5 demonstrates this visually by plotting the distribution of predicted risk scores under the old and new grid.

[24] An alternative approach is to use the average of the five discontinuities $(\xi_{E,4} + \xi_{F,9} + \xi_{G,14} + \xi_{H,19} + \xi_{I,9})/5$, which yields similar results.

is no longer statistically distinguishable from zero, and after 3 years, the estimates suggest no difference in incarceration rates at a given month.

The reduced-form effects on committing a new offense and reincarceration are shown in the light gray and black lines, respectively. There is a negative effect on the probability of reoffending that lasts at least 3 years after sentencing and does not increase afterward. The fact that differences

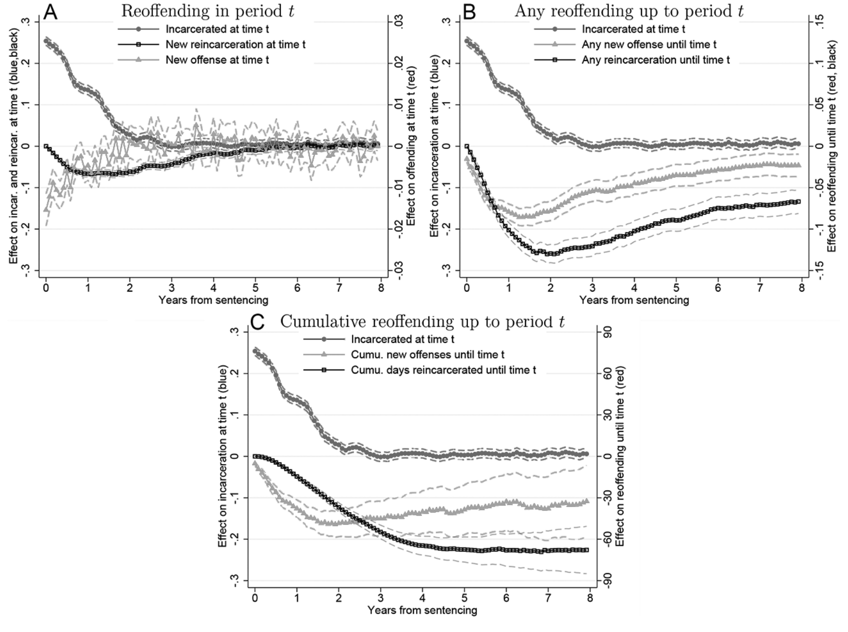

Fig. 4.—This figure shows the reduced-form effects of being to the right of a punishment type discontinuity on several key outcomes. The dark gray line with circles (left *Y*-axis) in all panels shows effects on an indicator for being incarcerated at the point in time on the *X*-axis. In *A*, the light gray line with triangles (right *Y*-axis) and black line with squares (left *Y*-axis) report effects on indicators for committing a new offense or being reincarcerated (i.e., after being released from any initial sentence) at time *t*, respectively. In *B*, the light gray and black lines report effects on indicators for ever committing a new offense and ever being reincarcerated up to time *t*, respectively. In *C*, the light gray line (right *Y*-axis) reports effects on the cumulative number of new offenses committed until *t*, with the black line reporting the cumulative number of days reincarcerated. We discretize time at the monthly level, so that there are 12 total estimates per year. Each point in each figure is an estimate of ξ^{RF}—an average of the reduced forms for each individual instrument—for the relevant outcome. This estimate is a constrained version of equation (1) that requires the coefficients of the five punishment type discontinuities to be equal. This strategy averages across all five offense classes and instruments but collapses our variation into a single coefficient (taking the actual average of the individual reduced forms yields highly similar results). Standard errors are clustered by individual. The regression specifications include as controls demographics (e.g., race, gender, age fixed effects), fixed effects for the duration of time previously incarcerated, the number of past incarceration spells and the number of past convictions, county fixed effects, and year fixed effects. Estimates without controls yield similar results (see table A.3). A color version of this figure is available online.

in offending rates stabilize at zero (or slightly below) is an indication that an initial term of incarceration does not increase criminal behavior in the long run. If it did, the light gray (and black) line would lie above zero.

Since within-month effects are noisily estimated relative to cumulative measures, such as ever committing a new offense, we next examine the reduced-form effects on any reoffending within *t* months from sentencing in figure 4*B*. This graph shows a permanent decrease in the probability of ever committing a new offense and even larger impacts on the likelihood of being reincarcerated. The decrease reaches a nadir after roughly 18–24 months, when the estimate begins to increase but remains negative 8 years after sentencing. This hook shape is what one would expect to see if individuals had a constant or decreasing hazard of reoffending after release and is not indicative of any criminogenic effects of incarceration. As initial incarceration sentences begin to expire, an increasing share of the treated group is released and has the opportunity to reoffend. By this point, however, many individuals not initially incarcerated have already reoffended, generating the slight increase after 18 months. The fact that new offenses and reincarceration stabilize below zero is again indicative that an initial term of incarceration does not increase criminal behavior in the long run. Effects on cumulative new offenses show a similar pattern, but effects stabilize earlier, after roughly 3 years, as shown in figure 4*C*.[25]

C. 2SLS Estimates

Table 2 reports 2SLS estimates using months of incarceration as the endogenous regressor and compares them with corresponding ordinary least squares (OLS) estimates with and without controls. We use any reincarceration within 5 years of sentencing as the outcome. While both OLS and 2SLS estimates are negative, 2SLS estimates are substantially more so, suggesting that unobserved selection is an important concern for OLS estimates. The 2SLS effects imply that 1 year of prison exposure reduces the likelihood of reincarceration within 5 years by 13.8 percentage points (↓28%). Reassuringly, 2SLS estimates change little when flexible controls for criminal history and demographics are included.

Table 3 reports 2SLS effects on alternative reoffending measures. The results show that incarceration generates substantial declines in reoffending across a broad set of offense types. While point estimates differ substantially,

[25] Following Bhuller et al. (2020), we also examine effects on the cumulative number of new offenses that occurred after 36 months, when the instruments no longer predict incarceration status in a given month (fig. 4). Any effects measured starting from 36 months after sentencing, therefore, cannot be attributed directly to mean differences in incapacitation. These estimates are relatively precise zeros (see fig. A.5), suggesting that incarceration does not have any criminogenic effects on reoffending between 3 and 8 years after sentencing.

TABLE 2
Effect of Incarceration on Reincarceration within 5 Years

	OLS (1)	OLS (2)	RD (3)	RD (4)
Months of incarceration	−.00651***	−.00847***	−.0115***	−.0123***
	(.0000373)	(.0000472)	(.000892)	(.000876)
1-year effect in percentages	−15.64	−20.35	−27.59	−29.43
Dependent variable mean among nonincarcerated	.500	.500	.500	.500
Controls	No	Yes	No	Yes
F statistic (excluded instruments)			154.9	155.6
Observations	451,547	451,547	451,547	451,547

Note.—This table presents OLS and 2SLS estimates of the effect of incarceration on an indicator for ever being reincarcerated within 5 years of the individual's sentencing date. Columns 1 and 2 show OLS estimates of eq. (2) using this outcome, while cols. 3 and 4 report 2SLS estimates using the five punishment type discontinuities as instruments. Controls include indicators for gender, age, race, ethnicity, number of previous cases, number of previous incarceration spells, months of previous incarceration, number of previous convictions, year of offense, county of conviction, and the offense code of the convicted offense. Standard errors (in parentheses) are clustered by individual. The F statistics test the joint hypothesis that the coefficients on the excluded instruments are all equal to zero. Because of clustering, the F statistic reported is cluster robust. Effective and nonrobust F statistics are similar. The number of observations is smaller than in table 1 because the sample in the regressions is restricted to individuals who are observed at least 5 years after the date of sentencing in our data. Standard errors are shown in parentheses.
*** $p < .001$.

effects are relatively similar when compared with the nonincarcerated means. One year of incarceration decreases the likelihood of committing any new offense by 13%, a new felony offense by 12%, a new violent offense by 22%, a new property offense by 11%, and a new drug offense by 6%. The largest reductions are for the more severe events, such as violent crime or reincarceration. Drug offenses are the only crime category for which there are no economically meaningful reductions 5 years after sentencing.[26]

Table A.8 reports 2SLS estimates splitting the sample by the category of the defendant's initial conviction. These results show that all types of offenders are affected by incarceration. While assault offenders are the main driver of the overall effects on new violent crime offenses, property and drug offenders also reduce offending across all categories of crime. Moreover, the effects persist even 8 years after sentencing. The estimates in table A.7 show that 1 year of incarceration causes a reduction of 8% in the likelihood of a new offense as well as an 8% reduction in the likelihood of a new felony offense and a 21% reduction in reincarceration.

[26] Table A.6 shows that effects are even larger when probation revocations are excluded by dropping offenders whose probation was revoked before committing a new offense. Specifically, 1 year of incarceration reduces the likelihood of a new offense within 5 years of sentencing by 18%, a new felony offense by 16%, a new violent offense by 29%, and reincarceration by 29%.

TABLE 3

EFFECT OF INCARCERATION ON ADDITIONAL REOFFENDING MEASURES WITHIN 5 YEARS

	Measure of Crime					
	Reincarceration (1)	Any New Offense (2)	Felony (3)	Violent (4)	Property (5)	Drug (6)
Months of incarceration	−.0123***	−.00554***	−.00359***	−.00235**	−.00185*	−.00111
	(.000876)	(.000880)	(.000881)	(.000743)	(.000753)	(.000726)
1-year effect in percentages	−29.43	−13.18	−11.59	−21.61	−11.33	−6.017
Dependent variable mean among nonincarcerated	.500	.504	.371	.131	.196	.221
F statistic (excluded instruments)	155.6	155.6	155.6	155.6	155.6	155.6
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	451,547	451,547	451,547	451,547	451,547	451,547

NOTE.—This table presents 2SLS estimates of the effect of incarceration on various outcomes. The dependent variable is an indicator for the event in the column header ever occurring within 5 years of sentencing. New offenses (both overall and by crime type) are measured using either arrests recorded in the AOC data or convictions recorded in the DPS data. We use the date at which the offense occurred rather than the date an individual was arrested or convicted. Controls include indicators for gender, age, race, ethnicity, number of previous cases, number of previous incarceration spells, months of previous incarceration, number of previous convictions, year of offense, county of conviction, and the offense code of the convicted offense. Standard errors (in parentheses) are clustered by individual. The F statistics test the joint hypothesis that the coefficients on the excluded instruments are all equal to zero. Because of clustering, the F statistic reported is cluster robust. Effective and nonrobust F statistics are similar. The number of observations is smaller than in table 1 because the sample in the regressions is restricted to individuals who are observed at least 5 years after the date of sentencing. Standard errors are shown in parentheses.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 4
Heterogeneity in the Effects on 5-Year Reincarceration Rates by Age and Previous Incarceration Exposure

	No Previous Incarceration (1)	Previous Incarceration (2)	≥28 (3)	<28 (4)
Months of incarceration	−.00978***	−.0127***	−.0115***	−.0118***
	(.00175)	(.00104)	(.00159)	(.00102)
1-year effect in percentages	−28.18	−22.61	−24.99	−31.60
Dependent variable mean among nonincarcerated	.416	.672	.553	.447
First-stage coefficient (incarceration length)	6.909	5.806	5.997	6.035
Controls	Yes	Yes	Yes	Yes
F statistic (excluded instruments)	42.82	112.9	47.37	116.7
Observations	247,530	204,017	216,552	234,995

Note.—This table shows heterogeneity in the effects of incarceration on an indicator for ever being reincarcerated within 5 years of the individual's sentencing date. Controls include indicators for gender, age, race, ethnicity, number of previous cases, number of previous incarceration spells, months of previous incarceration, number of previous convictions, year of offense, county of conviction, and the offense code of the convicted offense. Standard errors (in parentheses) are clustered by individual. The F statistics test the joint hypothesis that the coefficients on the excluded instruments are all equal to zero. Because of clustering, the F statistic reported is cluster robust. Effective and nonrobust F statistics are similar. The total number of observations is smaller than in table 1 because the sample in the regressions is restricted to individuals who are observed at least 5 years after the date of sentencing. Standard errors are shown in parentheses.

*** $p < .001$.

Table 4 reports 2SLS estimates by age and previous incarceration history. Overall, we find that incarceration causes a meaningful reduction in reoffending regardless of whether the offender previously spent time in prison. The effect of a year of incarceration is larger for individuals with previous exposure to prison (15.2 vs. 11.7 percentage points). However, because reoffending rates are much higher for individuals with a prior incarceration spell (67% vs. 42%), effects divided by baseline mean reoffending rates show a larger percentage reduction for individuals without a prior incarceration spell (↓28%) relative to offenders with previous prison experience (↓22%). Columns 3 and 4 in table 4 show that there is no meaningful heterogeneity in the effects of incarceration based on age—the 2SLS estimates are nearly identical. However, older offenders have higher recidivism rates (55% vs. 45%), making effects relative to baseline reoffending rates higher for younger offenders.[27]

[27] Figure A.6 reports reduced-form estimates by previous incarceration exposure for any reincarceration, any new offense, and cumulative measures of reoffending. The reduced-form effects for any reoffending are similar, but cumulative reoffending measures are larger for offenders without previous exposure to incarceration. Examining the reduced-form effects by age shows slightly larger reductions in reoffending among younger offenders (see fig. A.7).

Effect heterogeneity by felony class is discussed in appendix D. Overall, the patterns in each class look similar, although there is substantial variation in the shifts in incarceration exposure generated by each discontinuity. Estimates are also surprisingly stable over time. Table E.1 shows that incarceration length has a similar effect across offenders sentenced in different time periods (e.g., 1995–99 vs. 2010–14).[28]

D. 2SLS under Nonlinear and Heterogeneous Effects

The effects of incarceration may depend nonlinearly on the duration of exposure (the dose). For example, incarcerating an offender for 1 year may have a different impact than lengthening a 3-year sentence by another year. The effects may also vary with unobserved characteristics of the offender. For instance, a year of prison may have different impacts on individuals who would always be incarcerated under the current regime versus individuals who would never be. Unpacking how effects differ along both dimensions is critical for understanding the impact of potential changes in sentencing policy.

The 2SLS estimates presented above capture weighted averages of the effects of different doses of incarceration on different groups of compliers. Formally, let $Y_{it}(d)$ be an indicator for whether individual i would reoffend within t months after sentencing if incarcerated for d months. As before, D_i denotes months of incarceration and Z_i denotes whether the individual is above or below a punishment type discontinuity. Abstracting from pretreatment covariates, the 2SLS estimator with a single binary instrument recovers the average causal response (ACR) discussed in Angrist and Imbens (1995):

$$\frac{\mathbb{E}[Y_{it}|Z_i = 1] - \mathbb{E}[Y_{it}|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}\left[Y_{it}(d) - Y_{it}(d-1) \,\middle|\, D_i(1) \geq d > D_i(0)\right], \quad (3)$$

where $\omega_d = (\Pr(D_i(1) \geq d > D_i(0)))/(\sum_{l=1}^{\bar{D}} \Pr(D_i(1) \geq l > D_i(0)))$. Nonlinearity means that the dose response $\mathbb{E}[Y_{it}(d) - Y_{it}(d-1)]$ is a nonlinear function of d, while heterogeneity means that the dose response differs across individuals.[29] 2SLS estimates average across different doses for different populations, with weights and complier groups that depend on the instrument. When using multiple instruments, the 2SLS estimator

[28] For comparability with other studies in the literature, tables A.3, A.4, and A.5 report estimates analogous to those in tables 2, 3, and 4 but for 3-year reoffending/reincarceration measures rather than 5 years.

[29] The populations relevant to the ACR are defined by the conditions $D_i(1) \geq d > D_i(0)$. These are individuals who would be incarcerated for strictly fewer than d months when $Z_i = 0$ but otherwise would be incarcerated for at least d months.

captures a weighted average of instrument-specific ACRs.[30] Thus, it is unclear how the estimates relate to other important parameters, such as the ATE of a year of incarceration versus no prison time or the effects of marginal changes in sentencing policy.

Table 5 investigates the potential importance of nonlinearity and heterogeneity by using different instruments to identify the effect of an additional year of incarceration. If dose-responses are linear and homogeneous, estimated effects should be similar regardless of the instrument used, since the terms in the summation in equation (3) would all be the same. Column 1 uses all grid cell boundaries except those between the first and second prior record level, yielding 20 excluded instruments. Column 2 uses only the five punishment type discontinuities that shift offenders along both the extensive and the intensive margins, and column 3 uses the remaining 15, which primarily shift offenders along only the intensive margin at much longer durations. Interestingly, the effects in column 3 are meaningfully more negative than those in column 2, suggesting that the marginal impact of incarceration may be increasing in sentence length. Specifically, 1 year of prison causes a 13.8 percentage point reduction in the likelihood of reincarceration within 5 years in column 2 but a 19.8 percentage point reduction in column 3. Moreover, in both columns 2 and 3, overidentification tests reject the null hypothesis that the treatment effects identified by each of the five (or 15) instruments are the same. This suggests that the estimated effects would differ within the sets of instruments used in each column.

A simple way to allow for some nonlinearity in dose response is to include multiple endogenous variables that capture the effect of exposure to different amounts of incarceration. Column 4 does so by adding an indicator for any incarceration. Column 5 adds a quadratic term in incarceration length as well. These estimates show that marginal effects on the extensive margin effect (e.g., zero to one) are larger than those on the intensive margin effect (e.g., two to three). For example, in column 5, 1 year of incarceration reduces the likelihood of reincarceration by 30.5 percentage points starting from no exposure, but a shift from 2 to 3 (or 3 to 4) years reduces it by only 10.6 (or 3.5) percentage points.[31]

[30] Mogstad, Torgovitsky, and Walters (2020) discuss how 2SLS estimates based on multiple instruments can capture a linear combination of different treatment effects with potentially negative weights. In our setting, conditional on the required running variable controls, the set of instruments always takes on one of two distinct values (an individual cannot be just above two discontinuities at the same time). Imbens and Angrist (1994) monotonicity and partial monotonicity (Mogstad, Torgovitsky, and Walters 2020) therefore place the same restrictions on counterfactual treatment choices over the support of the instruments conditional on the covariates. The overidentified 2SLS specification in eqq. (1) and (2) therefore aggregates these covariate-specific, single-instrument 2SLS estimates into a weighted average with nonnegative weights.

[31] Table A.10 reports 2SLS estimates that treat probation revocations as random censoring by excluding from the sample observations with a revocation prior to any new criminal offense. In

The patterns of nonlinearity in columns 4 and 5 are the opposite of those suggested by comparing columns 2 and 3. Treatment effect heterogeneity is a likely explanation. The compliers shifted along the intensive margin in column 3 may have higher baseline recidivism propensities, explaining why marginal increases in incarceration generate large reductions in reoffending even at high doses. Such heterogeneity in effects across individuals could explain the differences between columns 2 and 3 even if the dose response function is linear for any given individual. Since 2SLS models with multiple endogenous variables do not have a local average treatment effect (LATE) interpretation, the estimates in columns 4 and 5 have a clear causal interpretation only under a constant treatment effect assumption. Thus, properly accounting for both nonlinearity and heterogeneity requires a more flexible framework and is one of the primary objectives of the selection model developed in section IV.[32]

Another way to assess the importance of treatment effect heterogeneity is to conduct the following out-of-sample prediction exercise. For any given instrument, we can combine estimates of its ACR weights in equation (3) and estimates of the dose response function from column 5 of table 5 to predict the estimated effect of months of incarceration one would obtain from a 2SLS procedure using that instrument alone. If treatment effect heterogeneity is unimportant, then this prediction should be close to the actual 2SLS estimate for that instrument. We construct predictions for the 15 primarily intensive margin grid discontinuities, which were not used to estimate the nonlinear 2SLS model in column 5. Figure A.10 reports the results. It is visually clear that the predicted effects do not accurately match the observed estimated effects. We show later, however, that using the selection model provides a more accurate replication.

E. Robustness Checks

1. Sorting through Plea Bargains

While prior record points are difficult to manipulate, plea bargains can affect the offense class in which an individual is ultimately convicted. Some offenders may thus be able to manipulate their vertical position in the sentencing grid. Although all individuals have incentives to plead down to

---

this sample, all reincarceration events are therefore the result of new criminal charges. The results are similar and display the same pattern of nonlinearity. We also estimate overidentification tests, which fail to reject in the multiple endogenous variable models. An important caveat to these tests is that they are necessary but not sufficient conditions for the presence of a nonlinear and heterogeneous dose-response function. For example, the test can fail to reject when treatment effect heterogeneity and nonlinearity cancel out in such a way as to make the ACR sufficiently similar regardless of the instrument.

[32] Note that for comparability with the selection model estimates, table 5 does not include any additional controls. Including additional controls yields very similar estimates, as shown in table A.9.

TABLE 5
Evidence for Nonlinearity and Heterogeneity in Treatment Effects

	Only Length of Incarceration			+Indicator for Any Sentence	+Polynomial Squared Term
	All (1)	Five Punishment Types (2)	15 Primarily Intensive (3)	Five Punishment Types (4)	Five Punishment Types (5)
Linear effects:					
0 to 1 year	−.144***	−.138***	−.198***		
	(.0104)	(.0107)	(.0260)		
Nonlinear effects:					
0 to 1 year				−.288***	−.305***
				(.0336)	(.0377)
1 to 2 years				−.0511*	−.106*
				(.0209)	(.0519)
2 to 3 years				−.0511*	−.0349
				(.0209)	(.0277)
3 to 4 years				−.0511*	.0366
				(.0209)	(.0801)
Dependent variable mean among nonincarcerated	.500	.500	.500	.500	.500
J statistic	65.62	25.33	30.33	2.392	.563
J statistic p-value	.0000005	.00004	.00687	.495	.755
Excluded instruments F statistics:					
Length of incarceration	44.31	154.9	12.70	100.7	2.560
Any incarceration				132.8	49.39
Length of incarceration^2					.871
Controls	No	No	No	No	No

Note.—This table shows the results of 2SLS regressions of the effect of incarceration on an indicator for ever being reincarcerated within 5 years of the individual's sentencing date. Each column shows the implied effect of increasing sentences by the amount indicated in the row from separate specifications. Columns 1–3 use our standard specification in eq. (1). Because the endogenous variable is simply months of prison, each effect is the same. Column 1 uses all 20 discontinuities as excluded instruments. Column 2 uses only the five punishment type discontinuities, as in our main results. Column 3 uses only the other 15 discontinuities. These instruments primarily shift sentences on the intensive margin. Column 4 augments this specification by adding a second endogenous variable, an indicator for any prison sentence. Column 5 then adds a third term for the squared length of the sentence. Both of these columns use only the five punishment type discontinuities as in col. 2. The J statistics and associated p-values refer to Sargan-Hansen tests of overidentifying restrictions. The tests examine whether the 2SLS estimates are consistent among different subsets of the instruments. Standard errors (in parentheses) are clustered by individual. We also report the F statistics for the excluded instruments for each of the different endogenous variables. In cols. 1–3, there is a single endogenous variable and the F statistics are all above the rule of thumb of 10 proposed by Stock, Wright, and Yogo (2002). In cols. 4 and 5, there are multiple endogenous variables, so we report the partial F statistic proposed by Angrist and Pischke (2009). Note that there are no clear rules of thumb regarding the size of the F statistic when there are multiple endogenous variables. The number of observations is smaller than in table 1 because the sample in the regressions is restricted to individuals who are observed at least 5 years after the date of sentencing. Standard errors are shown in parentheses.

* $p < .05$.

*** $p < .001$.

lesser charges, individuals whose initial charges put them just to the right of a large discontinuity in sentences may be especially incentivized to do so, since by pleading down to a lower offense class they can avoid any (or longer) incarceration sentences. Likewise, individuals may be less incentivized to plead to a charge that would result in a conviction just to the right of a major discontinuity, since the gains to doing so are smaller.

When defining our instruments using individuals' convicted charges, such sorting could potentially bias our estimates. To address this concern, we compare our primary estimates, which use the most severe convicted charge to define the instruments, with estimates that use the most severe charge at arraignment and most severe charge brought at any point in the case in table E.2. Arraigned offenses are determined at first appearance. Because law enforcement is the charging agency in North Carolina, these charges map very closely to actual arrested charges. In Charlotte-Mecklenburg County, where we collected arrest data directly from the sheriff, the charge on the arrest report matches the charge at arraignment in greater than 95% of cases. Thus, arraigned charges are unlikely to be affected by plea negotiation. Using the arraigned offense yields results very similar to using the convicted offense, confirming that plea-induced selection is not an issue. The main difference is that the standard errors on the estimates using the convicted charge are roughly 40% smaller. In appendix E.1, we discuss an additional test that compares the characteristics of individuals who take a plea with those who do not and also shows no evidence of manipulation through plea bargaining.

2. Differences in the Likelihood of Detection

Individuals on probation are supervised closely. Their criminal activity may be detected more often than that of offenders initially sentenced to incarceration. Our estimated effects of incarceration therefore may capture both differences in the propensity to commit crimes and differences in the likelihood of getting caught. To examine whether differences in the likelihood of detection are driving any of our results, we conduct two separate analyses. Overall, both pieces of evidence reveal that our estimated effects are likely not driven by difference in detection probabilities and instead reflect the causal effects of incarceration on criminal activity itself.

First, we show that our results remain the same when using only discontinuities that primarily shift the length of incarceration exposure rather than the margin of probation versus prison. These 15 discontinuities are the three other grid cell boundaries in each offense class besides the five punishment type discontinuities used in the majority of the analysis. Figure E.3 presents 2SLS estimates of the effects of incarceration length on reoffending within t months from sentencing using this variation. These estimates show patterns similar to those in our core reduced-form analysis in figure 4. The estimates also do not change when including probation revocations in the measure of reoffending.

Second, we exploit a discontinuity in the guidelines that shifts offenders from community punishment to intermediate punishment, both of which are probation regimes but with different levels of monitoring. In class I, when offenders move between prior record levels I and II, the recommended sentence changes from either community or intermediate punishment to only intermediate punishment. Figure E.4 documents the first-stage effects on the probation regime and shows that there is no effect on the likelihood of reoffending or being reincarcerated within 3 years of sentencing. Figure E.5 shows that this discontinuity has no effects on any pretreatment characteristics (e.g., race, age at offense, etc.). In addition, the likelihood of being sentenced to an active term of incarceration also does not change at the discontinuity.[33]

IV. Nonlinear and Heterogeneous Effects

Thus far, our analysis has studied the effects of incarceration on reoffending identified solely by our quasi-experimental variation. In this section, we introduce a single-index generalized Roy (1951)–style selection model that allows us to push beyond these results in several important ways. The model allows for treatment effects that are both potentially nonlinear in total exposure to incarceration and heterogeneous across individuals. Our previous analysis allowed for only one or the other but not both.[34] We use the model to bound treatment effects of clearly defined doses (e.g., 1 vs. 0 years of incarceration) for policy-relevant populations and the average offender, allowing us to clearly characterize incarceration's dose-response function. We then investigate the importance of unobserved heterogeneity and selection by examining how dose-response varies across observably equivalent offenders.

A. Model

Treatment is discrete and ordered, with $D_i \in \{0, \dots, \bar{D}\}$ (i.e., months incarcerated). There is one potential outcome for each level of exposure, and observed outcomes are given by $Y_i = \sum_{d=0}^{\bar{D}} \mathbf{1}\{D_i = d\} Y_i(d)$. Treatment is determined by the following set of selection equations:

[33] These findings are similar to others in the literature. Georgiou (2014), e.g., utilizes a salient discontinuity in the level and intensity of supervision in Washington State and also finds no effects on reoffending.

[34] Relatedly, albeit in a different context, previous work on the effects of welfare programs on labor supply found evidence of both nonlinearities in treatment effects (Kline and Tartari 2016) and heterogeneity (Bitler, Gelbach, and Hoynes 2006).

$$\mathbf{1}\{D_i \geq d\} = \mathbf{1}\left\{C^d(X_i, Z_i) - V_i^d \geq 0\right\} \text{ for } d \in \{1, \dots \bar{D}\}, \tag{4}$$

where V_i^d is a random variable and C^d represents unknown functions of observables X_i and instruments Z_i satisfying $C^{d-1}(X_i, Z_i) - V_i^{d-1} \geq C^d(X_i, Z_i) - V_i^d \quad \forall\ i, d$. One interpretation of $C^d(x, z)$ is as the perceived cost to judges of imposing a sentence of at least length d for offenders with observables and values of the instrument x and z, respectively. The sentencing guidelines affect these costs by changing legally permissible sentences (North Carolina General Statutes §15A–81B). The latent index $-V_i^d$ can thus be interpreted as judges' perceived benefits of sentencing offender i to at least d months. Offenders receive such a sentence whenever benefits outweigh the costs.[35]

We make the following standard exogeneity assumption:

Assumption 1 (Exogeneity). $\{V_i^d\}_{d=1}^{\bar{D}}, \{Y_i(d)\}_{d=1}^{\bar{D}} \perp\!\!\!\perp Z_i | X_i$.

Vytlacil (2006b) shows that under these assumptions, this model is equivalent to the extension of the LATE model for an ordered treatment maintained in the preceding analysis (Angrist and Imbens 1995).[36] We strengthen these assumptions further by assuming that a single latent factor V_i determines treatment rather than the full set $\{V_i^d\}_{d=1}^{\bar{D}}$:

Assumption 2 (Single latent factor). $V_i^d = V_i \quad \forall\ d \in \{1, \dots, \bar{D}\}$.

This assumption reduces the dimensionality of unobservables that may affect potential outcomes and is a version of the single-index restriction common in the program evaluation literature (e.g., Meghir and Palme 1999; Dahl 2002; Heckman, Urzua, and Vytlacil 2006; Heckman and Vytlacil 2007a, 2007b).[37] Assumption 2 also allows us to write the selection equation as a standard ordered-choice problem where the thresholds

[35] Under this interpretation, the condition on the ordering of $C^d(X_i, Z_i) - V_i^d$ therefore requires the net benefit of sentencing a given offender to at least d months to be weakly decreasing in d. This seems uncontroversial since sentences of at least length d nest all possible sentences of at least length $d + 1$.

[36] Vytlacil (2006b) considers a model with random thresholds where, conditional on X_i, treatment choice is determined by $\mathbf{1}\{D_i = d\} = 1\{\xi_i^{d-1} < \nu(Z_i) \leq \xi_i^d\}$, with $\xi_i^{d-1} \leq \xi_i^d$ for all i, d. In our notation, treatment choice is given by $\mathbf{1}\{D_i = d\} = 1\{C^{d+1}(Z_i) - V_i^{d+1} < 0 \leq C^d(Z_i) - V_i^d\}$. Hence, letting $\xi_i^d - v(Z_i) = C^d(Z_i) - V_i^d$ makes the two models equivalent. Clearly, when $\bar{D} = 1$, this model is also identical to the canonical program evaluation model (Heckman and Vytlacil 1999, 2005). Note that the monotonicity restriction in the LATE framework (Imbens and Angrist 1994) implies that the selection index is additively separable with respect to the latent factor, V_i^d (Vytlacil 2002, 2006a). In other words, monotonicity implies that $f^d(X_i, Z_i, V_i^{*d}) = C^d(X_i, Z_i) - V_i^d$ for any function $f(\cdot)$ and $V_i^d = g(V_i^{*d})$ for some function g.

[37] This assumption imposes cross-person restrictions on treatment assignment not imposed by Angrist and Imbens's (1995) monotonicity assumption. Specifically, the instruments cannot induce changes in potential treatment orderings across units. That is, for any i, j with $X_i = X_j$, $D_i(0) < D_j(0), D_i(1) > D_i(0), D_i(1) > D_j(0) \Rightarrow D_j(1) \geq D_i(1)$. For example, if $D_i(0) = 1, D_i(1) = 3$, and $D_j(0) = 2$, then $D_j(1) \geq 3$. Standard monotonicity requires only that $D_j(1) \geq D_j(0)$. We view this assumption as reasonable in our setting since changes in guideline sentences are unlikely to affect judges' ranking of how harshly to punish defendants.

depend on observable pretreatment covariates and excluded instruments. For expositional convenience, we assume that V_i is continuously distributed conditional on X_i and impose a standard normalization using its cumulative distribution function, $F_{V|X}$. Treatment assignment is then given by

$$\begin{aligned}\mathbf{1}\{D = d\} &= \mathbf{1}\left\{F_{V|X}(C^{d+1}(X_i, Z_i)) \leq F_{V|X}(V_i) < F_{V|X}(C^{d}(X_i, Z_i))\right\} \\ &= \mathbf{1}\{\pi_{d+1}(X_i, Z_i) \leq U_i < \pi_d(X_i, Z_i)\},\end{aligned} \tag{5}$$

where $U_i \sim$ Uniform $[0, 1]$ conditional on X_i. The random variable $\pi_d(X_i, Z_i) = \Pr(D_i \geq d|X_i, Z_i)$ has the natural interpretation of an ordered-choice propensity score, with $\Pr(D_i = d|Z_i, X_i) = \pi_d(X_i, Z_i) - \pi_{d+1}(X_i, Z_i)$. The random variable U_i represents the standard unobserved resistance to treatment. Individuals with lower U_i are assigned longer sentences and vice versa.

Average potential outcomes under each treatment dose d are functions of U_i and X_i:

$$E[Y_i(d)|U_i = u, X_i = x] = m_d(u, x),$$

where $m_d(u, x)$ represents the marginal treatment response (MTR) function (Mogstad, Santos, and Torgovitsky 2018). When combined, MTRs define the set of possible treatment effects across different doses of incarceration conditional on U_i and X_i:

$$\mathrm{MTE}_{d',d}(u, x) = E[Y_i(d') - Y_i(d)|U_i = u, X_i = x] = m_{d'}(u, x) - m_d(u, x).$$

These marginal treatment effect (MTE) functions (Heckman and Vytlacil 1999, 2005) measure the causal effect of a change in incarceration exposure from d to d' for any fixed value of u and x. Integrating over u and x therefore gives mean treatment effects for relevant populations. For example, the mean treatment effect of d' versus d units of incarceration for individuals with observables x, z and assigned incarceration dose k is given by $\int_{\pi_{k+1}(x,z)}^{\pi_k(x,z)} \mathrm{MTE}_{d',d}(u, x)\, du/(\pi_k(x, z) - \pi_{k+1}(x, z))$.

Importantly, the model so far makes no assumptions about why individuals are more or less resistant to treatment or how this selection process relates to outcomes. It may be the case, for example, that outcomes such as reoffending are unrelated to treatment assignment conditional on X_i, implying that $m_d(u, x) = m_d(x)$. In this case, simple OLS regressions of reoffending on d conditional on X_i would recover ATEs for this population. Alternatively, individuals more likely to be sentenced to longer prison terms (i.e., low U_i) may also be unobservably more likely to reoffend. This selection pattern is consistent with the differences between our OLS and 2SLS estimates in the preceding analysis.

Our goal, therefore, is to learn about the unknown functions m_d and to thereby characterize nonlinearity and heterogeneity in the treatment effects of incarceration. To do so, one option would be to make parametric assumptions on m_d that allow it to be point identified. For example, the classic Heckit approach assumes that $V_i|X_i \sim N(0, 1)$ and that MTRs are linear in V_i, implying that $m_d(u, x) = \alpha(x, d) + \beta(x, d)\Phi^{-1}(u)$ (Heckman 1974, 1976, 1979).[38] Garen (1984) and Card's (1999) selection model for ordered treatments implies that MTRs are linear in u conditional on X_i and d: $m_d(u, x) = \alpha(x, d) + \beta(x, d)u$. More generally, MTE models defined by a fixed number of parameters can be identified with sufficient exogenous variation in treatment propensity (Moffitt 2008; Brinch, Mogstad, and Wiswall 2017). Fully nonparametric identification requires instruments that generate continuous support over the probability of treatment (Heckman and Vytlacil 1999, 2001a, 2005).

Alternatively, one can use instruments to partially identify parameters of interest and therefore characterize certain features of the unknown treatment response functions. Sharp bounds for parameters such as the ATE (among others) have been studied in a wide variety of settings (e.g., Manski 1989, 1990; Balke and Pearl 1997; Haile and Tamer 2003; Shaikh and Vytlacil 2011). In recent work, Mogstad, Santos, and Torgovitsky (2018) lay out a general framework for bounding policy-relevant treatment effects (Heckman and Vytlacil 2001b, 2005; Carneiro, Heckman, and Vytlacil 2010). This approach provides a computationally tractable method that also easily allows the researcher to incorporate shape restrictions, such as monotonicity and monotone treatment response (Manski 1997; Manski and Pepper 2000, 2009), derived from economic theory.

We extend Mogstad, Santos, and Torgovitsky's (2018) approach to the ordered treatment case and provide bounds under various assumptions. To do so, we approximate $m_d(u, x)$ using a flexible function of u and x. We then find m_d functions that minimize or maximize the parameter of interest under two restrictions. First, each m_d must be consistent with the data. Specifically, it must reproduce the quasi-experimental variation induced by the instruments. Second, each m_d must satisfy shape restrictions motivated by theory or our specific context. Depending on the flexibility of the approximation to m_d and the impact of shape restrictions, this approach may point identify the target parameter or yield bounds.

As a leading example, consider the ATE of d versus $d - 1$ units of incarceration (conditional on X_i). This parameter corresponds to

$$\text{ATE}_{d,d-1}(X_i) = \int_0^1 m_d(u, X_i)\, du - \int_0^1 m_{d-1}(u, X_i)\, du.$$

[38] Note, however, that in many cases two-step control function estimators of Heckit models yield LATE estimates that are numerically equivalent to those produced by instrumental variable estimators (Kline and Walters 2019).

Our analysis estimates lower and upper bounds for $\text{ATE}_{d,d-1}(X_i)$. We do so by picking a set of $m_d(u, x)$ functions that minimize and maximize it, respectively. We follow Mogstad, Santos, and Torgovitsky (2018) and model $m_d(u, x)$ using Bernstein polynomials, which provide a scalable degree of flexibility and important analytical advantages, as discussed in appendix F.

To be consistent with the data, candidate m_d must reproduce the means of Y_i conditional on D_i, Z_i, and X_i:

$$\mathbb{E}[Y_i | D_i = d, Z_i = z, X_i = x] = \frac{1}{\pi_d(x, z) - \pi_{d+1}(x, z)} \int_{\pi_{d+1}(x,z)}^{\pi_d(x,z)} m_d(u, x)\, du. \quad (6)$$

Forcing candidate MTRs to match these moments implies that the MTRs can also reproduce single endogenous variable 2SLS estimates and complier means from the 2SLS analysis in the above section, since all such objects are linear combinations of these moments and the π's. In this sense, the conditional means exhaust all the available information about outcomes in the data (see app. F.1).

We then further restrict each m_d to satisfy various shape restrictions that help discipline the relationship between unobservables and outcomes. For example, in much of what follows we impose that $\partial m_d(x, u)/\partial u \leq 0$, a version of monotone treatment selection (Manski and Pepper 2000, 2009). This restriction implies that individuals whom judges would otherwise sentence to more prison time are more likely to be reincarcerated conditional on receiving a given sentence d.[39] In section IV.C.3, we estimate bounds on the selection process directly and find strong empirical evidence in support of this assumption.

A second important class of shape restrictions that we consider involves the separability of observable factors X_i and unobservables U_i. For example, a common assumption imposes additive separability by specifying $m_d(u, x) = f_d(x) + g_d(u)$, which implies that selection on unobservables works the same way for every value of the covariates and allows the researcher to use variation in X_i to help pin down g_d (e.g., Carneiro, Heckman, and Vytlacil 2011; Kline and Walters 2016; Brinch, Mogstad, and Wiswall 2017; Bhuller et al. 2018). In our setting, the most critical observables are individuals' locations in North Carolina's sentencing grid (i.e., prior points and felony class). Additive separability thus restricts differences in unobserved treatment effect heterogeneity for individuals with different criminal histories and convicted of different offenses.

In constructing our bounds, we focus only on the observables necessary for our research design—individuals' prior points and felony class—and

[39] Blundell et al. (2007) use an analogous shape restriction on participation in the labor market. They assume that individuals with higher wages are more likely to be employed.

omit all others. Since our 2SLS estimates apply to individuals at each sentencing discontinuity, to avoid additional extrapolation we consider values of X_i that fall exactly at each punishment type discontinuity.[40] This leaves five values of X_i, one for each felony class. Our baseline case allows for unrestricted differences across MTR functions at each value of X_i. Our analysis therefore bounds target parameters for each class separately or bounds the average of treatment effects associated with each.

We also estimate additively separable models as discussed above, allowing for treatment effect heterogeneity based on observables and unobservables but not their interaction. Finally, we estimate bounds imposing a stronger restriction that limits the impact of observables to a single coefficient across all MTRs, or $m_d(u, x) = f(x) + g_d(u)$, essentially imposing similarity of treatment effects (i.e, MTEs) across X_i but not across outcome levels or u. As discussed below, both constraints provide substantial identifying power but also represent nontrivial restrictions on the data-generating process. When implementing them, we first assess their plausibility by testing whether they are consistent with observed data using goodness-of-fit tests developed by Chernozhukov, Newey, and Santos (2020).

B. *Estimation*

Estimation proceeds in three main steps. We provide an overview of the method here, leaving full details for appendix F.

1. Estimate π_d for each d, Z_i, and X_i.
2. Estimate the conditional means $\mathbb{E}[Y_i|D_i, Z_i, X_i]$ for each value of D_i, Z_i, and X_i.
3. Bound target parameters subject to constraints and shape restrictions.

To accomplish step 1, we estimate equation (5) using an ordered probit model. Each threshold $C^d(Z_i, X_i)$ depends on Z_i and X_i using the same specification as in the reduced-form analysis, the right-hand side of equation (1).[41] We estimate $\pi_d(X_i, Z_i)$ as the fitted probabilities that $D_i \geq d$ for

[40] In the notation of our primary reduced-form specification (eq. [2]), $X_i = [p_i, \text{class}_i]'$. Z_i is an indicator for whether an individual falls to the right or left of the punishment type discontinuity in her class, or $1\{p_i \geq l_k\}1\{\text{class}_i = k\}$ for each $k \in$ classes and class-specific prior points threshold l_k.

[41] Allowing the thresholds to depend on Z_i and X_i can be thought of as flexibly modeling the variation in incarceration spells that is introduced by the nonlinearities in the guidelines. Other studies using ordered-choice models with thresholds that depend on covariates (or are themselves random variables) include Cameron and Heckman (1998), Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2007), and Greene and Hensher (2010).

the values of X_i at each discontinuity and $Z_i \in \{0, 1\}$.[42] Intuitively, these fits measure the probability of receiving a sentence of at least length d just to the left and just to the right of each discontinuity. For our five discontinuities and $\bar{D}$ doses of incarceration, this yields $5 \cdot \bar{D} \cdot 2$ total π_d's. In our base case, we discretize treatment into 3-month doses of incarceration.[43]

To accomplish step 2, one would ideally estimate equation (1) but with Y_i as the outcome and the sample restricted to observations with $D_i = d$ for each value of d. Conditional means would then be taken from the fitted values just to the left and just to the right of each discontinuity. In practice, we find that these estimates can be quite noisy, making it difficult to conduct step 3. We therefore use conditional means taken from an estimate of equation (1) that adds interactions of all covariates with a third-order polynomial in d, uses Y_i as the outcome, and includes only observations in grid cells adjacent to each discontinuity. Estimated conditional means are the fits for each value of d, for the values of X_i at each discontinuity, and for $Z_i \in \{0, 1\}$.

We then proceed to step 3. Given our choice to approximate the MTRs using Bernstein polynomials, bounds on interesting treatment effects can be computed as the solution to a linear programming problem. Figure 5 provides a graphical illustration of how restrictions on MTRs affect the bounds. This figure plots the MTRs consistent with the minimum and maximum ATE of receiving 12 months versus no prison on reincarceration within 5 years of sentencing. The MTRs plotted are those for individuals with values of X_i at the punishment type discontinuity in class I. The figure also plots the conditional means of the outcome for individuals at the same discontinuity who actually receive 12 and 0 months of incarceration and with $Z_i = 0$ and $Z_i = 1$. The bars at the bottom of the graph plot the range of u for the individuals who contribute to these means (the bounds of the integral in eq. [6]). The ATE is simply the area between the MTRs for 0 and 12 months over the full support of u.

In figure 5*A*, MTRs are restricted to be Bernstein polynomials of degree five, to fall between zero and one, and to match the estimated conditional means. This requires, for example, that the area under the MTR for $d = 0$ over the range shown by the bar labeled $d0$, $z0$ be equal to the conditional mean labeled $d0$, $z0$. The ATE bounds in figure 5*A* are wide: −53 to 21 percentage points. This is unsurprising, since without additional assumptions the MTRs are free to take very extreme shapes while remaining consistent with the empirical moments.

[42] For example, for the punishment type discontinuity in class I, we take the fits with $\text{class}_i = 9$ and $p_i = 8.5$ to get π_d when $Z_i = 1$.

[43] We have also explored more granular units, such as 1-month intervals. Doing so trades off the accuracy of our estimates of the π's and conditional means, which would rely on less data, against allowing for different treatment effects across finer doses. We view 3 months as striking an appropriate balance.

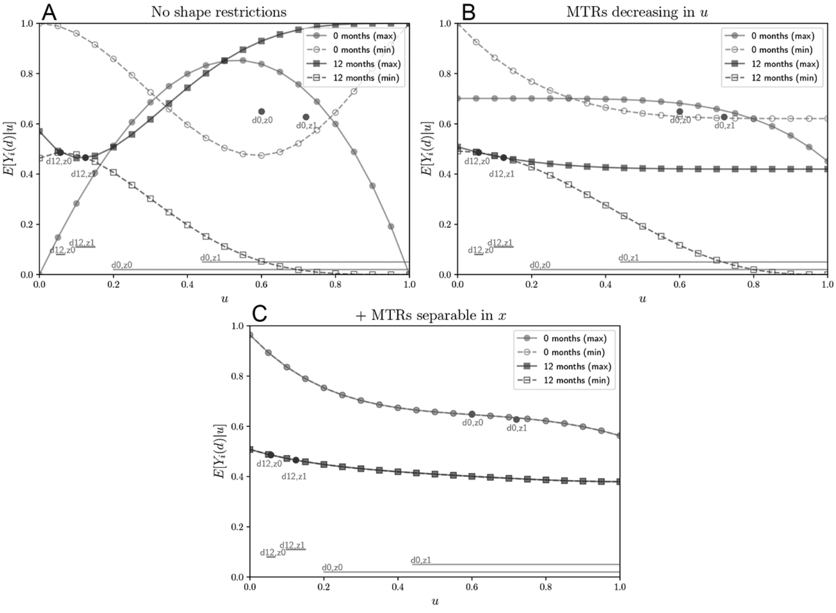

FIG. 5.—This figure illustrates estimation of ATE bounds under varying shape restrictions. Each panel plots MTRs underlying bounds of the ATE of increasing sentences from 0 to 12 months for offenders at the class I punishment type discontinuity. The outcome is an indicator for reincarceration within 5 years of sentencing. MTRs are all estimated using a Bernstein polynomial of degree five. Each panel also plots the conditional outcome means for individuals with $d \in \{0, 12\}$ months and $Z_i \in \{0, 1\}$ (the dots) and the ranges of u for individuals who contribute to these means (the bars along the X-axis). In A, MTRs are constrained to fall between zero and one to match these conditional means when integrated over the indicated range of u. Panel B adds an additional shape restriction that MTRs are weakly increasing in $-u$, implying that individuals who are more likely to receive longer sentences are more likely to be reincarcerated conditional on a sentence d. Panel C adds the assumption that MTRs take the form $m_d(x, u) = f_d(x) + g_d(u)$ and hence are additively separable in observables and unobservables. Additive separability allows variation across x to help pin down the shape of $g_d(u)$. With the addition of this final assumption, the ATE is nearly point identified. A color version of this figure is available online.

Figure 5B adds the restriction that MTRs are decreasing in u. As discussed above, this implies that individuals whom judges would otherwise sentence to more prison time are more likely to be reincarcerated conditional on a given sentence d. It does not require that judges consider only recidivism risk when making incarceration decisions. If judges consider other factors, however, they cannot lead to more risky individuals receiving shorter sentences on average. Under this simple restriction, the bounds are surprisingly informative. The ATE is between -48 and -23 percentage points. Our previous 2SLS estimates in table 5 are either at the top of this range, when we impose linear effects of incarceration in column 1, or closer to the bottom, when we allow for nonlinearities but shut down any treatment effect heterogeneity in column 5. These bounds therefore demonstrate that when allowing for both nonlinear and heterogeneous effects, exposure to incarceration generates a large decline in the likelihood of being incarcerated over a 5-year horizon for the average offender.

In figure 5*C*, we add a final shape restriction requiring MTRs to be additively separable in the impact of observables and unobservables—in other words, that $m_d(x, u) = f_d(x) + g_d(u)$—for the values of x at the five punishment type discontinuities. This assumption allows all instruments to contribute to identification of each $g_d(u)$, providing substantial identifying power relative to the nonseparable models in figure 5*A* and 5*B*. In fact, the ATE in figure 5*C* is nearly point identified and equal to about −28 percentage points. While separability assumptions are common in the program evaluation literature, they place meaningful restrictions on the data-generating process by ruling out treatment effect heterogeneity in the interaction between x and u. Nevertheless, our goodness-of-fit tests below show that we cannot reject that the data are consistent with separable MTRs.

C. *Results*

1. ATE Estimates

Table 6 presents estimates of ATEs of incarceration for offenders with values of the running variable that place them at the punishment type discontinuity in each class. The only shape constraint we impose in this analysis is that the MTR functions are weakly decreasing in u, implying that the unobserved factors that lead offenders to be sentenced to longer incarceration terms are weakly positively correlated with their propensity to reoffend. As in our previous results, the outcome is an indicator for any reincarceration within 5 years of sentencing. We use a Bernstein polynomial of degree five to approximate the MTRs. Using more flexible approximations changes results little, as we discuss below.

The first three rows bound the ATEs of incarcerating an offender for an additional year. Since we make no assumptions about the separability of observables and unobservables in treatment effects, each felony class has its own bounds. Across all classes, however, the estimates point to reductions in reincarceration rates that are largest in the first year of exposure. For example, in class I, which consists of the least severe felonies, such as breaking into a vehicle or possessing small quantities of cocaine, the ATE for 1 year versus no incarceration is between −48 and −22 percentage points. ATEs for 1 versus 2 years and 2 versus 3 years are much closer to zero (in fact, the bounds include meaningful positive treatment effects, although no other classes do so). Similarly, the ATE of 0 versus 1 year of incarceration in class E, which includes more severe offenses,

TABLE 6
BOUNDS ON ATEs OF INCARCERATION

	Class I (1)	Class H (2)	Class G (3)	Class F (4)	Class E (5)	Average (6)	Average and Separate MTRs (7)	Average and Same MTEs (8)
Marginal effects:								
0 to 1 year	[−.48, −.22]	[−.34, −.26]	[−.35, −.14]	[−.31, −.12]	[−.30, −.04]	[−.40, −.18]	[−.23, −.23]	[−.22, −.22]
1 to 2 years	[−.18, .17]	[−.13, −.13]	[−.33, −.09]	[−.28, −.13]	[−.18, −.09]	[−.20, .02]	[−.09, −.09]	[−.11, −.11]
2 to 3 years	[−.22, .08]	[−.22, −.11]	[−.26, −.04]	[−.12, −.00]	[−.08, −.07]	[−.19, .01]	[−.09, −.07]	[−.13, −.13]
3 to 4 years	[−.24, .04]	[−.26, −.16]	[−.23, −.12]	[−.14, −.06]	[−.23, −.14]	[−.22, −.05]	[−.20, −.18]	[−.16, −.16]
Total effects:								
0 to 2 years	[−.44, −.27]	[−.47, −.39]	[−.55, −.36]	[−.52, −.31]	[−.40, −.22]	[−.46, −.30]	[−.32, −.32]	[−.33, −.33]
0 to 3 years	[−.53, −.33]	[−.69, −.50]	[−.70, −.50]	[−.56, −.39]	[−.48, −.30]	[−.56, −.38]	[−.40, −.38]	[−.46, −.46]
Goodness-of-fit *p*-value						.938	.528	.236

NOTE.—This table presents bounds on the ATE of various doses of incarceration. The outcome is an indicator for any reincarceration within 5 years of sentencing. Each bound is the minimum or maximum value of the ATE associated with all possible MTR functions that (*a*) rationalize the quasi-experimental moments generated by our instruments and (*b*) satisfy certain shape constraints. MTRs are approximated using Bernstein polynomials of degree five and are constrained to be decreasing in u, the unobserved resistance to treatment. Each bound corresponds to the marginal or total effect listed in the row. MTRs take the form $m_d(x, u)$, where x includes prior points and felony class only. We consider the five distinct values of x that would place an offender exactly at the punishment type discontinuity in each class. In cols. 1–6, we impose no restriction on the relationship between unobserved and observed heterogeneity by fitting separate MTR functions for each value of x, yielding five different sets of bounds. Column 6 bounds the average of effects across each discontinuity, weighted by the sample frequency of offenders in adjacent grid cells. Column 7 adds the constraint that $m_d(x, u) = f_d(x) + g_d(u)$, allowing for heterogeneity in observed and unobserved treatment effects but not their interaction. Column 8 strengthens this assumption by requiring $m_d(x, u) = f(x) + g_d(u)$, implying the same MTEs at each u for each value of x. Note that bounds on marginal effects do not sum to bounds on total effects because the MTR functions overlap between marginal effects (e.g., 0 to 1 year and 1 to 2 years both depend on the MTR for 1 year of incarceration), implying that the lower bounds across marginal effects are not necessarily consistent. See sec. IV for additional details on the approach. The final row of the table reports p-values from goodness-of-fit tests analogous to a J-test in an overidentified GMM. The null hypothesis is that the data-generating process both satisfies the imposed shape constraints and fits the reduced-form moments. The p-value in col. 7, e.g., tests whether MTRs that are bounded between zero and one, decreasing in u, and satisfy separability in x and u are consistent with the data. To conduct inference, we use the procedure proposed by Chernozhukov, Newey, and Santos (2020) for shape constraint GMM settings (for details, see app. F).

such as second-degree kidnapping, is a reduction between 4 and 30 percentage points. Exposure to additional years of incarceration generates potentially smaller but still economically meaningful decreases in the likelihood of reincarceration.

The final two rows estimate bounds on ATEs for the total effects of shifting an offender from no prison time to 2 or 3 years of incarceration. These bounds are informative and include large and negative (i.e., crime-reducing) treatment effects.[44] They are also remarkably consistent across felony classes (especially the effects for 2 years), suggesting that observable factors such as prior points and felony class may not play a large role in treatment effect heterogeneity for certain total effects. Column 6 reports bounds on the average of treatment effects across the five discontinuities. These bounds show a pattern similar to class-specific estimates, with crime-reducing treatment effects that are largest in the first year.

The final two columns of table 6 impose shape restrictions on MTRs across felony classes. Column 7 imposes separability, so that $m_d(x, u) = f_d(x) + g_d(u)$. Column 8 strengthens this assumption by requiring $m_d(x, u) = f(x) + g_d(u)$, which implies that MTEs for any dose of incarceration are the same for each value of x. Both assumptions imply that the same $g_d(u)$ functions must rationalize more quasi-experimental moments, providing substantial identifying variation. The bound on the ATE of 1 versus 0 years of incarceration in columns 7 and 8 is now point identified and equal to roughly -23 percentage points. As in the felony class-specific estimates, reductions are largest for the first year of incarceration in both separable models.[45]

The final row of table 6 reports p-values from goodness-of-fit tests for some models. These tests are analogous to a J-test in an overidentified generalized method of moments (GMM) problem. The null hypothesis is that the data are generated by MTRs that satisfy the imposed shape constraints (e.g., additive separability) and fit the reduced-form moments in equation (6). To conduct inference, we use Chernozhukov, Newey, and Santos's (2020) goodness-of-fit test for shape-constrained GMM (for full details, see app. F). The goodness-of-fit tests show that we cannot reject that MTRs are decreasing in u or additive separability between the observed and unobserved factors in the MTR function (col. 7). We also cannot reject the stronger restriction that the MTEs are the same across the five felony classes (col. 8).

[44] Because each marginal effect bound is calculated in isolation, they do not sum to bounds on total effects. The ATE for 2 vs. 1 years is $\mathbb{E}[Y_i(2) - Y_i(1)|X_i = x]$, while the ATE for 1 vs. 0 years is $\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]$. Hence, if $\mathbb{E}[Y_i(1)]$ is more negative, as in the lower bound for the effect of 1 vs. 0 years, the ATE for 2 vs. 1 years must be more positive.

[45] Table A.13 reports bound estimates when treating probation revocations as random censoring by dropping individuals who had a probation revocation before committing any new offense. These estimates show patterns similar to the ones in table 6, although nonlinearities are less pronounced when imposing separability assumptions.

Estimates from column 8 allow us to use the model to predict treatment effects out of sample, since the effect of observables $f(x)$ are differenced off in any MTE. We use this restricted model to construct bounds on the 2SLS estimated obtained using each of the 15 intensive-margin discontinuities one at a time. These discontinuities were not used to construct our primary 2SLS estimates or to construct moments targeted in our bounding procedure. Figure A.10 shows that, compared with multiple endogenous variable 2SLS estimates, the model provides a meaningfully tighter fit to the observed 2SLS estimates for these discontinuities.

Approximating MTRs using Bernstein polynomials of degree five provides a large amount of flexibility relative to standard control function estimators. For example, a classic Heckit approach would allow mean outcomes to depend on unobservables with one parameter for each discrete dose of incarceration. Using Bernstein polynomials of degree five allows outcomes to depend on unobservables with six separate parameters for each discrete dose. Estimates of the average ATE across discontinuities using higher-degree approximations are presented in table A.12. These estimates are very similar even when using a Bernstein polynomial of degree 15 or 20. Thus, we view these ATE estimates as embedding meaningfully fewer assumptions than alternative approaches.

To asses the uncertainty associated with our estimates, table A.11 uses the numerical bootstrap procedure proposed by Hong and Li (2020) to estimate pointwise valid 90% confidence intervals for the treatment effects in table 6.[46] As expected, intervals are wider than our point estimates but show similar patterns. The total effect of 2, 3, and 4 years of incarceration remains unambiguously crime reducing. The estimates also still point to larger reductions in reincarceration due to the first year of exposure than from further exposure.

2. Effects of Incarceration on Cumulative Reoffending Measures

Next, we examine the effects of incarceration on cumulative measures of reoffending. To understand what types of crimes are averted by prison, we also estimate bounds on the effects on different types of offenses (e.g., violent, drug, property). Table 7 reports bounds on the average effects (ATEs) across the five discontinuities for several margins. All bounds are estimated assuming only that MTR functions are weakly decreasing in u and using a Bernstein polynomial of degree five. Cumulative days reincarcerated have a natural upper bound of 5 years minus the initial sentence (e.g., 4 years

[46] For additional details on the implementation of the numerical bootstrap, see app. F. Confidence intervals that are uniformly consistent (in the data-generating process) are more challenging to construct in this setting (Shea and Torgovitsky 2020).

TABLE 7

BOUNDS ON ATEs FOR CUMULATIVE REOFFENDING MEASURES

	Days Reincarcerated (1)	New Offenses or Revocations (2)	Violent (3)	Property (4)	Drugs (5)	Revocations (6)	Other Offenses (7)
Marginal effects:							
0 to 1 year	[−503.98, −117.14]	[−5.00, −.42]	[−.84, .04]	[−2.00, −.02]	[−.58, −.02]	[−.86, −.55]	[−.72, .13]
1 to 2 years	[−119.67, −22.75]	[−1.92, −.73]	[−.19, .06]	[−.52, −.13]	[−.59, −.40]	[−.16, −.08]	[−.47, −.18]
2 to 3 years	[−90.67, −70.34]	[−.84, −.37]	[−.25, −.08]	[−.16, −.13]	[−.33, −.16]	[.03, .09]	[−.13, −.09]
3 to 4 years	[−55.04, −37.77]	[−1.08, −.76]	[−.15, −.05]	[−.46, −.39]	[−.20, −.18]	[−.10, −.03]	[−.16, −.12]
Total effects:							
0 to 2 years	[−533.87, −229.67]	[−6.04, −2.03]	[−.87, −.04]	[−2.13, −.54]	[−1.16, −.43]	[−.94, −.71]	[−.93, −.31]
0 to 3 years	[−617.40, −307.15]	[−6.55, −2.71]	[−1.02, −.23]	[−2.28, −.67]	[−1.31, −.76]	[−.90, −.62]	[−1.03, −.42]

NOTE.—This table reports bounds on the ATE of varying doses of incarceration for different cumulative measures of reoffending within 5 years of sentencing. The outcome in col. 1 is cumulative days reincarcerated (i.e., excluding the initial sentence). The outcome in col. 2 is the cumulative new offenses (arrests recorded in the AOC data and convictions recorded in the DPS data) and probation revocations (recorded in the DPS data). The offenses considered in cols. 3–7 are subsets of the outcomes included in col. 2. Each bound is the minimum or maximum value of the ATE associated with all possible MTR functions that (*a*) rationalize the quasi-experimental moments generated by our instruments and (*b*) satisfy certain shape constraints. MTRs are approximated using Bernstein polynomials of degree five and are constrained to be decreasing in u, the unobserved resistance to treatment. Each bound corresponds to the marginal or total effect listed in the row. MTRs take the form $m_d(x, u)$, where x includes prior points and felony class only. We consider the five distinct values of x that would place an offender exactly at the punishment type discontinuity in each class. Reported bounds reflect bounds on the average over x with weights proportional to the same frequency of offenders in adjacent prior record levels. In col. 1, MTRs are bounded above by 5 years minus the initial sentence (e.g., 4 years for a 1-year sentence MTR), the natural upper bound for days reincarcerated at a 5-year horizon. In cols. 2–7, MTRs are bounded above by the 95th percentile of each outcome. Note that bounds on marginal effects do not sum to bounds on total effects because the MTR functions overlap between marginal effects (e.g., 0 to 1 year and 1 to 2 years both depend on the MTR for 1 year of incarceration), implying that the lower bounds across marginal effects are not necessarily consistent. See sec. IV for additional details on the approach.

for an offender who served a 1-year sentence). Since the other cumulative reoffending measures are in principle unbounded, we impose an upper bound on MTR functions corresponding to the 95th percentile of each outcome.

Column 1 shows that a year of incarceration causes economically meaningful reductions in incarceration in the future. For example, the first year of incarceration causes an average reduction of between 117 and 504 days spent in prison (excluding the initial sentence) within 5 years of sentencing. This implies that the net cost of a year in prison is less than one-third the nominal price. A transition from 2 to 3 years causes smaller but still meaningful reductions of between 70 and 91 days.

Incarceration also causes a reduction in reoffending across the different crime categories, as is clearly seen in the total effects of 2 or 3 years of incarceration reported at the bottom of the table. Nonlinear effects are also evident across the different crime categories. For example, the first year of incarceration reduces between 0 and 0.84 violent crime arrests in the 5-year period after sentencing. Lengthening a 2-year sentence by an additional year reduces the number of violent crime arrests by between 0.08 and 0.25. The effects on probation revocations are especially concentrated in the first year of prison. This happens because only offenders not sentenced to prison—and hence placed on probation—are at risk for revocation.

3. Selection Patterns

Next, we examine how incarceration sentences relate to individuals' unobserved propensities to reoffend and their potential responses to time in prison. These selection patterns provide important evidence on the allocative efficiency of the justice system with respect to public safety, as well as insight into how reoffending risk impacts judges' sentencing decisions.

We begin by examining how expected reoffending rates under no incarceration, $\mathbb{E}[Y_i(0)|D_i = d]$, vary with assigned sentences or selection on levels. If $\mathbb{E}[Y_i(0)|D_i = d]$ is increasing in d, then individuals who are more likely to reoffend are sentenced to longer incarceration spells and thus $\partial m_0(x, u)/\partial u \leq 0$. Figure 6*A* shows that this is indeed the case: the unobserved factors that lead judges to mete out a longer prison sentence are strongly related to individuals' likelihood of reoffending. Figure 6*A* makes no assumptions about how MTRs vary with respect to u but does impose additive separability of x and u in the MTR function as in column 7 of table 6. The increasing pattern in figure 6*A* clearly supports the assumption that MTR functions are weakly decreasing in u maintained elsewhere.

Figure 6*A* shows that those most likely to reoffend are given the longest prison sentences. But are these individuals also the most likely to benefit (in terms of reduced recidivism) from time in prison? Figure 6*B* examines

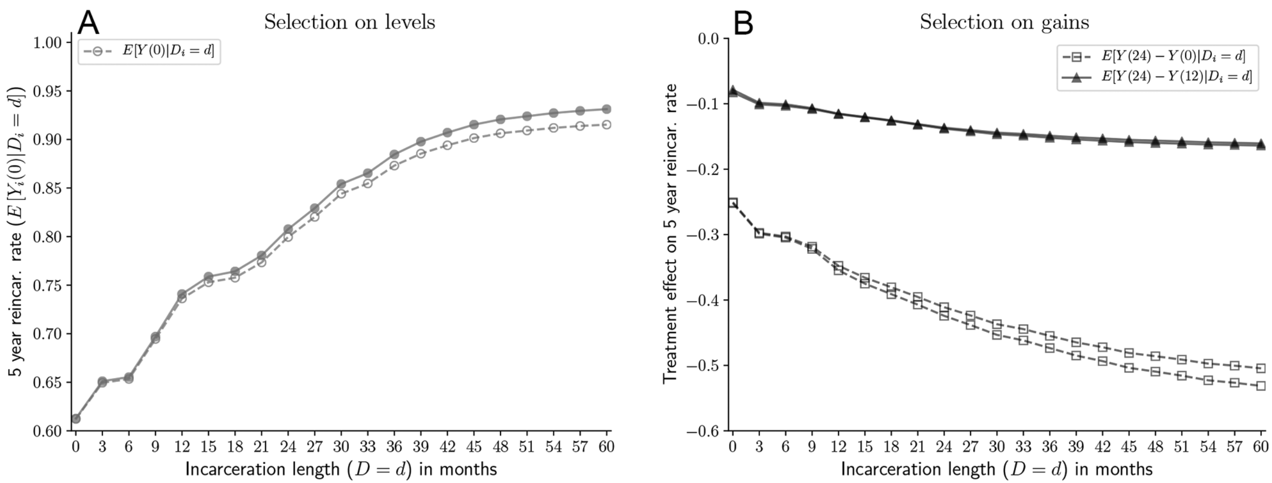

FIG. 6.—This figure plots estimates of bounds on the selection patterns in assignment to incarceration. Panel *A* plots bounds on selection into incarceration on levels (i.e., $\mathbb{E}[Y_i(0)|D_i = d]$) for different values of incarceration lengths *d*. This measures how much individuals who are sentenced to longer incarceration terms are also more likely to reoffend. Panel *B* plots bounds on selection into incarceration based on gains, $\mathbb{E}[Y_i(d) - Y_i(0)|D_i = d]$. This estimand reflects the degree to which individuals sentenced to longer incarceration spells experience larger/smaller changes in reoffending rates due to prison. For example, if $\mathbb{E}[Y_i(d) - Y_i(0)|D_i = 0] < \mathbb{E}[Y_i(d) - Y_i(0)|D_i > 0]$, this implies that incarcerated offenders have a larger treatment effect from incarceration. MTRs are approximated using Bernstein polynomials of degree five. Both panels constrain the MTRs to take the form $m_d(x, u) = f_d(x) + g_d(u)$—in other words, that MTRs are additively separable in observables *x* and unobserved heterogeneity *u*. In *A*, there are no other constraints. In *B*, MTRs are constrained to also be monotonically decreasing in *u*. This restriction implies that individuals whom judges would otherwise sentence to more prison time are more likely to be reincarcerated conditional on receiving a given sentence *d*. This assumption is supported by the results in panel *A*. A color version of this figure is available online.

this question by bounding selection on gains, $\mathbb{E}[Y_i(d'') - Y(d')|D_i = d]$. This object is the treatment effect of increasing incarceration exposure from d' to d'' for individuals currently sentenced to d months of prison. We examine two types of selection on gains in figure 6*B*: (i) a transition from no prison to 2 years of incarceration (i.e., $\mathbb{E}[Y_i(24) - Y(0)|D_i = d]$) and (ii) a purely intensive margin shift from 1 to 2 years of prison (i.e., $\mathbb{E}[Y_i(24) - Y(12)|D_i = d]$). A decreasing pattern is consistent with judges sentencing offenders for whom incarceration will reduce reoffending the most to longer prison terms.

Figure 6*B* shows clear evidence of selection on gains. The treatment effects become more negative for higher values of d, with effects that are meaningfully larger in magnitude for incarcerated versus nonincarcerated (i.e., $d = 0$) offenders. While the treatment effect slope is negative for both $\mathbb{E}[Y_i(24) - Y(0)|D_i = d]$ and $\mathbb{E}[Y_i(24) - Y(12)|D_i = d]$, it is much steeper for $\mathbb{E}[Y_i(24) - Y(0)|D_i = d]$, indicating that there is stronger selection on gains along the extensive margin. Thus, individuals assigned to prison experience larger reductions in reoffending from their exposure. This finding holds for individuals shifted along both the extensive and the intensive margin.

The findings in figure 6 indicate that treatment effects on individuals sent to prison under current policy should be larger than on the average offender. Table A.14 reports estimates of the average treatment effects on the treated (TOTs) of incarceration rather than ATEs—that is, $\mathbb{E}[Y_i(d'') - Y_i(d')|D_i = d']$ versus $\mathbb{E}[Y_i(d'') - Y_i(d')]$. Indeed, the TOT effects are larger for all doses of incarceration. Moreover, the nonlinearity that was documented for ATEs is also present for TOTs. Table A.15 reports TOT estimates for cumulative reoffending of different types. These TOTs also show larger reductions in reoffending than corresponding ATEs and exhibit a similar pattern of nonlinearity.

V. Policy Implications

In this section, we investigate some of the policy implications of the previous results. We begin by using the selection model and bounding framework developed in section IV to conduct several policy counterfactuals. Specifically, we bound the effects of a series of feasible and budget-neutral changes in sentencing that reduce the length of incarceration sentences overall in exchange for increasing the share of offenders given any prison sentence. Since we find that incarceration's impacts are largest for the initial exposure, such reallocations might reduce average reoffending rates. However, since we also find evidence of selection into incarceration based on gains, the full impact is ambiguous.

Figure 7 summarizes the results of this exercise, which we conduct separately for each discontinuity. The *X*-axis measures the share of offenders

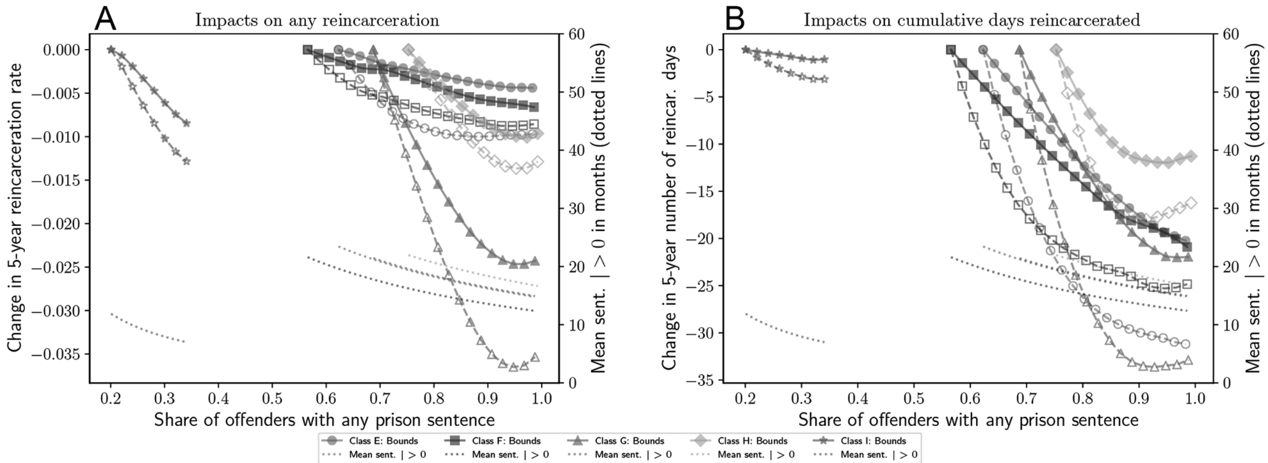

FIG. 7.—This figure reports the results of budget-neutral counterfactual exercises that reduce longer prison sentences and use the additional resources to incarcerate more offenders for short prison sentences. In each counterfactual, we reduce average sentence length among those sent to prison (dotted lines labeled "Mean sent. | > 0," measured on the right *Y*-axis) and increase the share of offenders sent to a short prison sentence (*X*-axis) in each offense class while holding total incarceration spending constant. Lines demarcated with symbols bound the impact on 5-year reincarceration. Leftmost points, where the estimated impact is zero, reflect current sentencing policy in each offense class. Bounds stop when it is no longer feasible to continue budget-neutral reallocations—for example, because 100% of offenders are imprisoned. In *A*, the outcome is any reincarceration event within 5 years of sentencing. In *B*, the outcome is the cumulative number of days spent reincarcerated (not including the original sentence) within 5 years of sentencing. MTRs are approximated using Bernstein polynomials of degree five. The only shape constraint we impose is that MTRs are monotonically decreasing in u (i.e., offenders who are sentenced to longer incarceration spells have higher baseline recidivism rates). A color version of this figure is available online.

given any prison sentence (i.e., $\Pr(D_i > 0)$). The right Y-axis measures the mean sentence length among those sent to prison (i.e., $\mathbb{E}[D_i|D_i > 0]$). In figure 7*A*, the left Y-axis measures the reduction in 5-year reincarceration rates as the mean sentence length is reduced and the share of offenders receiving any sentence increases. Figure 7*B* reports the reductions in the cumulative number of days an offender is reincarcerated within 5 years (i.e., not including the initial sentence). The bounds all begin at zero impact, reflecting the current sentencing regime. Because the sentencing guidelines differ at each discontinuity, the status quo involves different incarceration rates and average sentence lengths for each offense class. Moving to the right, we trace out bounds on the impact of trading off the intensive and extensive sentencing margin.

These estimates are produced by shifting estimates of the relevant π_d's in each offense class (i.e., X_i) and for $Z = 0$.[47] Given our discretization, $d = 0$ reflects no prison time, $d = 1$ reflects 3 months, and so on. We do not change π_d's for $d \geq 20$, since these thresholds affect offenders incarcerated for the full horizon over which we measure outcomes. To reduce mean sentence lengths, we shift π_{19} toward π_{20} incrementally. To increase the share serving any prison sentence, we shift π_1 toward one. Shifting both by the same amount is always budget neutral. When $\pi_{19} = \pi_{20}$, we shift π_{18}, then π_{17}, and so on. We stop when we reach π_1 or when $\pi_1 = 1$, implying that no more budget-neutral reallocations of this sort are possible. The end result is that all sentences are pulled toward the smallest unit—3 months.

Figure 7 shows that across all five discontinuities these shifts always reduce recidivism and in many cases can reduce it significantly. In classes E and G, for example, incarcerating nearly all offenders but cutting mean sentence lengths by 50% reduces cumulative days reincarcerated by as much as 1 month per offender. The rate of reincarceration also decreases because of such budget-neutral shifts in prison resources. The large variation in impacts across classes is at least partially due to differences in the initial distribution of sentences. For example, in class I only a small fraction of individuals face sentences of more than 2 years, providing limited scope for gains from reducing sentence length.[48]

Figure 7 imposes no restrictions on MTRs across classes. Figure A.9 reports the impacts when MTR functions are additively separable in x and u. This shape restriction tightens the bounds considerably. The increase in

[47] One could also easily study effects relative to the regime with $Z = 1$ or an average of the two. Reductions in reoffending are qualitatively similar but smaller at $Z = 1$, since there are fewer offenders currently receiving no prison sentence to be shifted along the extensive margin.

[48] Figure A.8 shows the distribution of offenders across incarceration sentences just below each of the discontinuities. There is large variation in the proportion of offenders who are not incarcerated. For example, in class H it is 25%, while in class E it is 43%, and in class I it is 80%. There is also variation across the discontinuities in the prevalence of longer sentences.

precision allows us to compare the magnitude of the gains as the reallocations become more extreme. For example, in class F, the reductions in recidivism are decreasing and plateau when the share of incarcerated individuals approaches one. This pattern is consistent with the selection on levels and gains documented earlier. As the reallocation becomes more extreme, the marginal individual shifted from zero to some prison sentence has a smaller treatment effect. Interestingly, for cumulative days reincarcerated, all the felony classes exhibit no indication of plateauing.

Estimating the effects of similar counterfactuals with 2SLS estimates instead of the selection model would yield meaningfully different conclusions. The linear effects suggested by the simple single endogenous variable model, for example, imply no gains to reallocation. Any effects of reducing exposure to long sentences are exactly offset by the effects of increasing short sentences. 2SLS specifications that allow for nonlinearities (e.g., col. 5 of table 5) rule out any unobserved heterogeneity. These models substantially overstate the potential gains from reallocations since they do not take into account selection on gains.

Finally, we describe how the estimates in section IV.C.2 can inform a partial cost-benefit analysis. Broadly, the economic benefit of a term of incarceration can be summarized by the following effects:

$$\begin{aligned} \text{net benefit} = {} & \Delta(\text{new crimes}) \cdot (\text{average \$ cost of crime}) \\ & + [\text{duration of sentence} + \Delta(\text{future incarceration})] \\ & \cdot (\text{\$ cost of incarceration}) \\ & + \text{general deterence effects} + \text{spillovers}. \end{aligned} \tag{7}$$

The estimates in table A.15 can be used to construct simple comparisons of the gains from reoffending reductions relative to the costs of incarceration. The first year of incarceration has an average net cost of between $(365 - 254) \cdot \$103 = \$11{,}433$ and $(365 - 188) \cdot \$103 = \$18{,}231$.[49] To justify these costs, the average value of an averted new offense or probation revocation needs to be between $\$11{,}433/1.23 = \$9{,}295$ and $\$18{,}231/0.78 = \$23{,}373$. This can be thought of as the break-even valuation of each criminal arrest needed to justify the costs of the first year of incarceration. Similar estimates can be derived for other margins. Generally, break-even valuations for the first year of prison are meaningfully lower than in later years. For example, lengthening a 2-year sentence by an additional year requires valuing the average averted offense by roughly $36,000.

[49] The cost estimate comes from the North Carolina Department of Public Safety (https://www.ncdps.gov/adult-corrections/cost-of-corrections) and provides cost estimates as of June 2019.

As with counterfactuals, using 2SLS estimates (from table A.16) to perform cost-benefit analysis yields meaningfully different conclusions. The 2SLS estimates understate the effects of incarceration on cumulative reoffending. For example, according to the selection model bounds, the ATE of 2 years of incarceration on any reoffending (col. 2 of table 7) is a reduction of two to six new offenses (treatment effects on the treated, as shown in table A.15, are between one and 15). The 2SLS model, however, estimates a reduction of only 0.46 new offenses from a year of incarceration.

VI. Concluding Remarks

Our analysis shows that incarceration substantially reduces reoffending in the years after sentencing. The effects are not concentrated among a specific type of criminal incident: we observe reductions in violent, property, and drug crimes, as well as reincarceration overall. We then use a Roy-style selection model to parse the heterogeneous dose response underlying these effects. We find that the ATEs of incarceration are diminishing in sentence length. In addition, we find that while the offenders given longest sentences have the highest recidivism risk, they also experience the largest reductions in reoffending due to exposure to prison.

Budget-neutral changes in sentencing that take advantage of these patterns by shortening sentences overall but sending a larger fraction of offenders to prison can generate meaningful reductions in recidivism. This exercise, however, speaks only to better allocations of sentences given current levels of incarceration spending; it does not imply that all offenders should be incarcerated for at least a brief period. Indeed, a broader cost-benefit analysis may find that it is optimal to reduce incarceration overall.

Our estimates are an important contribution to the ongoing debate over US criminal justice policy. After growing steadily since the 1970s, incarceration rates began to decline slightly in the mid-2000s. Recent policy changes, however, have the potential to at least check these recent reductions.[50] While our estimates show that incarceration sentences do not increase reoffending, they also demonstrate that incarceration has room to rehabilitate inmates further, especially when compared with carceral regimes in other developed countries, such as Norway. Since incarceration is unlikely to be abolished in the near future, understanding what features of imprisonment itself can be rehabilitative or damaging to offenders is a useful area for future research.

[50] See, e.g., Attorney General Jeff Sessions's reversal of the so-called Holder memo mitigating the impact of mandatory minimum sentences for drug crimes: http://www.politico.com/story/2017/05/12/mandatory-minimum-drug-sentences-jeff-sessions-238295.

References

Agan, Amanda, and Michael Makowsky. 2021. "The Minimum Wage, EITC, and Criminal Recidivism." *J. Human Resources* 56 (3): 1220–11398R1.

Agan, Amanda, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *Q.J.E.* 133 (1): 191–235.

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *Q.J.E.* 130 (2): 759–803.

Aneja, Abhay, and Carlos Fernando Avenancio-Leon. 2020. "No Credit for Time Served? Incarceration and Credit-Driven Crime Cycles." Unpublished manuscript.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *J. American Statis. Assoc.* 90 (430): 431–42.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton Univ. Press.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Q.J.E.* 133 (4): 1885–932.

Arteaga, Carolina. 2020. "Parental Incarceration and Children's Educational Attainment." Working paper, Dept. Econ., Univ. Toronto.

Austin, James, and Robert Lawson. 1998. "Assessment of California Parole Violations and Recommended Intermediate Programs and Policies." Tech. report, Nat. Council Crime and Delinquency, San Francisco.

Balke, Alexander, and Judea Pearl. 1997. "Bounds on Treatment Effects from Studies with Imperfect Compliance." *J. American Statis. Assoc.* 92 (439): 1171–76.

Barbarino, Alessandro, and Giovanni Mastrobuoni. 2014. "The Incapacitation Effect of Incarceration: Evidence from Several Italian Collective Pardons." *American Econ. J.: Econ. Policy* 6 (1): 1–37.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen. 2009. "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections." *Q.J.E.* 124 (1): 105–47.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *J.P.E.* 76 (2): 169–217.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2018. "Intergenerational Effects of Incarceration." *AEA Papers and Proc.* 108:234–40.

———. 2020. "Incarceration, Recidivism, and Employment." *J.P.E.* 128 (4): 1269–324.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *A.E.R.* 96 (4): 988–1012.

Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. 2007. "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." *Econometrica* 75 (2): 323–63.

Bowers, Jake, and Ben B. Hansen. 2009. "Attributing Effects to a Cluster Randomized Get-Out-the-Vote Campaign." *J. American Statis. Assoc.* 104 (487): 873–85.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. 2017. "Beyond LATE with a Discrete Instrument." *J.P.E.* 125 (4): 985–1039.

Buonanno, Paolo, and Steven Raphael. 2013. "Incarceration and Incapacitation: Evidence from the 2006 Italian Collective Pardon." *A.E.R.* 103 (6): 2437–65.

Cameron, Stephen V., and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *J.P.E.* 106 (2): 262–333.

Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, vol. 3A, edited by Orley Ashenfelter and David Card, 1801–63. Amsterdam: North-Holland.

Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *Internat. Econ. Rev.* 44 (2): 361–422.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2010. "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin." *Econometrica* 78 (1): 377–94.

———. 2011. "Estimating Marginal Returns to Education." *A.E.R.* 101 (6): 2754–81.

Chen, M. Keith, and Jesse M. Shapiro. 2007. "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-Based Approach." *American Law and Econ. Rev.* 9 (1): 1–29.

Chernozhukov, Victor, Whitney K. Newey, and Andres Santos. 2020. "Constrained Conditional Moment Restriction Models." Unpublished manuscript.

Clark, Damon, and Emilia Del Bono. 2016. "The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom." *American Econ. J.: Appl. Econ.* 8 (1): 150–76.

Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2007. "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds." *Internat. Econ. Rev.* 48 (4): 1273–309.

Dahl, Gordon B. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica* 70 (6): 2367–420.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *A.E.R.* 108 (2): 201–40.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. 2018. "The Intergenerational Effects of Parental Incarceration." Working Paper no. 24186, NBER, Cambridge, MA.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova. 2009. "The Deterrent Effects of Prison: Evidence from a Natural Experiment." *J.P.E.* 117 (2): 257–80.

Estelle, Sarah M., and David C. Phillips. 2018. "Smart Sentencing Guidelines: The Effect of Marginal Policy Changes on Recidivism." *J. Public Econ.* 164:270–93.

Franco, Catalina, David J. Harding, Jeffrey Morenoff, and Shawn D. Bushway. 2020. "Failing to Follow the Rules: Can Imprisonment Lead to More Imprisonment without More Actual Crime?" Unpublished manuscript.

Ganong, Peter N. 2012. "Criminal Rehabilitation, Incapacitation, and Aging." *American Law and Econ. Rev.* 14 (2): 391–424.

Garen, John. 1984. "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable." *Econometrica* 52 (5): 1199–218.

Georgiou, Georgios. 2014. "Does Increased Post-release Supervision of Criminal Offenders Reduce Recidivism? Evidence from a Statewide Quasi-experiment." *Internat. Rev. Law and Econ.* 37:221–43.

Green, Donald P., and Daniel Winik. 2010. "Using Random Judge Assignments to Estimate the Effects of Incarceration and Probation on Recidivism among Drug Offenders." *Criminology* 48 (2): 357–87.

Greene, William, and David Hensher. 2010. *Modeling Ordered Choices*. Cambridge: Cambridge Univ. Press.

Grogger, Jeffrey. 1995. "The Effect of Arrests on the Employment and Earnings of Young Men." *Q.J.E.* 110 (1): 51–71.

Haile, Philip A., and Elie Tamer. 2003. "Inference with an Incomplete Model of English Auctions." *J.P.E.* 111 (1): 1–51.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway. 2017. "Short- and Long-Term Effects of Imprisonment on Future Felony Convictions and Prison Admissions." *Proc. Nat. Acad. Sci. USA* 114 (42): 11103–8.

———. 2018. "Imprisonment and Labor Market Outcomes: Evidence from a Natural Experiment." *American J. Soc.* 124 (1): 49–110.

Heckman, James J. 1974. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42 (4): 679–94.

———. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." In *Annals of Economic and Social Measurement*, vol. 5, no. 4, edited by Sanford V. Berg, 475–92. Cambridge, MA: NBER.

———. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.

Heckman, James J., and Edward J. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proc. Nat. Acad. Sci. USA* 96 (8): 4730–34.

———. 2001a. "Local Instrumental Variables in Nonlinear Statistical Inferences." In *Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, edited by Cheng Hsiao, Kimio Morimune, and James L. Powell, 1–46. New York: Cambridge Univ. Press.

———. 2001b. "Policy-Relevant Treatment Effects." *A.E.R.* 91 (2): 107–11.

———. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73 (3): 669–738.

———. 2007a. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman and Edward E. Leamer, 4779–874. Amsterdam: North-Holland.

———. 2007b. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman and Edward E. Leamer, 4875–5143. Amsterdam: North-Holland.

Heckman, James J., Sergio Urzua, and Edward J. Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Rev. Econ. and Statis.* 88 (3): 389–432.

Hjalmarsson, Randi. 2009. "Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?" *J. Law and Econ.* 52 (4): 779–809.

Hong, Han, and Jessie Li. 2020. "The Numerical Bootstrap." *Ann. Statis.* 48 (1): 397–412.

Huttunen, Kristiina, Martti Kaila, Tuomas Kosonen, and Emily Nix. 2019. "Shared Punishment? The Impact of Incarcerating Fathers on Child Outcomes." Unpublished manuscript.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Kessler, Daniel, and Steven D. Levitt. 1999. "Using Sentence Enhancements to Distinguish between Deterrence and Incapacitation." *J. Law and Econ.* 42 (S1): 343–64.

Kline, Patrick, and Melissa Tartari. 2016. "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach." *A.E.R.* 106 (4): 972–1014.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Q.J.E.* 131 (4): 1795–848.
———. 2019. "On Heckits, LATE, and Numerical Equivalence." *Econometrica* 87 (2): 677–96.
Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *A.E.R.* 96 (3): 863–76.
Kuziemko, Ilyana. 2013. "How Should Inmates Be Released from Prison? An Assessment of Parole versus Fixed-Sentence Regimes." *Q.J.E.* 128 (1): 371–424.
Kyckelhahn, Tracey. 2011. "Justice Expenditures and Employment, FY 1982–2007 Statistical Tables." Report no. NCJ 236218, Bur. Justice Statis., US Dept. Justice, Washington, DC.
Lee, David S. 2008. "Randomized Experiments from Non-random Selection in US House Elections." *J. Econometrics* 142 (2): 675–97.
Levitt, Steven D. 1996. "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation." *Q.J.E.* 111 (2): 319–51.
———. 1998. "Juvenile Crime and Punishment." *J.P.E.* 106 (6): 1156–85.
Loeffer, Charles E. 2013. "Does Imprisonment Alter Life Course? Evidence on Crime and Employment from a Natural Experiment." *Criminology* 51 (1): 137–66.
Lofstrom, Magnus, and Steven Raphael. 2016. "Crime, the Criminal Justice System, and Socioeconomic Inequality." *J. Econ. Perspectives* 30 (2): 103–26.
Londoño-Vélez, Juliana, Catherine Rodríguez, and Fabio Sánchez. 2020. "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia." *American Econ. J.: Econ. Policy* 12 (2): 193–227.
Looney, Adam, and Nicholas Turner. 2018. "Work and Opportunity Before and After Incarceration." Report, Brookings Inst., Washington, DC.
Manski, Charles F. 1989. "Anatomy of the Selection Problem." *J. Human Res.* 24 (3): 343–60.
———. 1990. "Nonparametric Bounds on Treatment Effects." *A.E.R.* 80 (2): 319–23.
———. 1997. "Monotone Treatment Response." *Econometrica* 65 (6): 1311–34.
Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68 (4): 997–1010.
———. 2009. "More on Monotone Instrumental Variables." *Econometrics J.* 12 (S1): S200–S216.
Marvell, Thomas B., and Carlisle E. Moody. 1994. "Prison Population Growth and Crime Reduction." *J. Quantitative Criminology* 10 (2): 109–40.
Maurin, Eric, and Aurelie Ouss. 2009. "Sentence Reductions and Recidivism: Lessons from the Bastille Day Quasi Experiment." IZA Discussion Paper no. 3990, Inst. Labor Econ., Bonn.
McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *J. Econometrics* 142 (2): 698–714.
McCrary, Justin, and Sarath Sanga. 2012. "General Equilibrium Effects of Prison on Crime: Evidence from International Comparisons." In *Cato Papers on Public Policy*, vol. 2, edited by Jeffrey Miron, 165–93. Washington, DC: Cato Inst.
Meghir, Costas, and Marten Palme. 1999. "Assessing the Effect of Schooling on Earnings Using a Social Experiment." IFS Working Paper no. W99/10, Inst. Fiscal Studies, London.
Miles, Thomas J., and Jens Ludwig. 2007. "The Silence of the Lambdas: Deterring Incapacitation Research." *J. Quantitative Criminology* 23 (4): 287–301.

Moffitt, Robert. 2008. "Estimating Marginal Treatment Effects in Heterogeneous Populations." *Ann. Econ. et Statis.* 91:239–61.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters." *Econometrica* 86 (5): 1589–619.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R. Walters. 2020. "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables." Working Paper no. 25691, NBER, Cambridge, MA.

Mueller-Smith, Michael. 2015. "The Criminal and Labor Market Impacts of Incarceration." Working paper, Dept. Econ., Univ. Michigan, Ann Arbor.

Nagin, Daniel S., and G. Matthew Snodgrass. 2013. "The Effect of Incarceration on Re-offending: Evidence from a Natural Experiment in Pennsylvania." *J. Quantitative Criminology* 29 (4): 601–42.

National Center for State Courts. 2008. "State Sentencing Guidelines Profiles and Continuum." Tech. report, Nat. Center State Courts, Williamsburg, VA.

Norris, Samuel. 2018. "Judicial Errors: Evidence from Refugee Appeals." Working Paper no. 2018-75, Becker Friedman Inst., Chicago.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2020. "The Effects of Parental and Sibling Incarceration: Evidence from Ohio." Working paper.

Owens, Emily G. 2009. "More Time, Less Crime? Estimating the Incapacitative Effects of Sentence Enhancements." *J. Law and Econ.* 52 (3): 551–79.

Raphael, Steven. 2014. *The New Scarlet Letter? Negotiating the U.S. Labor Market with a Criminal Record.* Kalamazoo, MI: W.E. Upjohn Inst.

Raphael, Steven, and Magnus Lofstrom. 2016. "Incarceration and Crime: Evidence from California's Realignment Sentencing Reform." *Ann. American Acad. Polit. and Soc. Sci.* 664 (1): 196–220.

Rose, Evan K. 2021. "Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders." *Q.J.E.* 136 (2): 1199–253.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Econ. Papers* 3 (2): 135–46.

Shaikh, Azeem M., and Edward J. Vytlacil. 2011. "Partial Identification in Triangular Systems of Equations with Binary Dependent Variables." *Econometrica* 79 (3): 949–55.

Shea, Joshua, and Alexander Torgovitsky. 2020. "ivmte: An R Package for Implementing Marginal Treatment Effect Methods." Working Paper no. 2020-01, Becker Friedman Inst., Chicago.

Solis, Alex. 2017. "Credit Access and College Enrollment." *J.P.E.* 125 (2): 562–622.

Stevenson, Megan. 2016. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." *J. Law, Econ., and Org.* 34 (4): 511–42.

———. 2017. "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails." *Rev. Econ. and Statis.* 99 (5): 824–38.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *J. Bus. and Econ. Statis.* 20:518–29.

US Department of Justice. 1996. "National Assessment of Structured Sentencing." Report, Bur. Justice Assistance, Washington, DC.

Vytlacil, Edward J. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica* 70 (1): 331–41.

———. 2006a. "A Note on Additive Separability and Latent Index Models of Binary Choice: Representation Results." *Oxford Bull. Econ. and Statis.* 68 (4): 515–18.

———. 2006b. “Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results.” *Rev. Econ. and Statis.* 88 (3): 578–81.

Yang, Crystal S. 2017. “Local Labor Markets and Criminal Recidivism.” *J. Public Econ.* 147:16–29.

Zapryanova, Mariyana. 2020. “The Effects of Time in Prison and Time on Parole on Recidivism.” *J. Law and Econ.* 63 (4): 699–727.