Does Incarceration Increase Crime?*

Evan K. Rose (UC Berkeley)

Yotam Shem-Tov (UC Berkeley)

May 29, 2019

Abstract

This paper studies the causal effect of incarceration on reoffending using discontinuities in sentencing guidelines and two decades of administrative records from North Carolina. A regression discontinuity analysis shows that one year of incarceration reduces the likelihood of committing new assault, property, and drug offenses within three years of conviction by 38%, 24%, and 20%, respectively. Incarceration sentences temporarily incapacitate offenders by removing them from society but can also influence post-release criminal behavior. To parse the non-linear and heterogeneous effects of these channels, we develop an econometric model of sentencing length and recidivism. Our model allows for Roy-style selection into sentencing on the basis of latent criminality. We propose a two-step control function estimator of the model parameters and show that our estimates accurately reproduce the reduced form effects of the sentencing discontinuities we study. Our parameter estimates indicate that incarceration has modest crime-reducing behavioral effects that are diminishing in incarceration length. A cost-benefit analysis suggests, however, that the benefit of reducing crime by lengthening sentences (through both incapacitation and behavioral channels) is outweighed by the large fiscal costs of incarceration.

*Yotam Shem-Tov (corresponding author): Postdoctoral Fellow at UC Berkeley and assistant Professor at the Economics Department of UCLA from July 2020, shemtov@berkeley.edu; Evan K. Rose: Ph.D. Candidate, Department of Economics, ekrose@berkeley.edu; 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 Avi Feller, Robert Gregory, Hilary Hoynes, Gabriel Lenz, Nicholas Li, Juliana Londoño-Vélez, Justin McCrary, Conrad Miller, Allison Nichols, Emmanuel Saez, 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 UC San Diego, UC Los Angeles, University of Michigan at Ann Arbor, Chicago Crime Lab, Society of Labor Economics Annual Meeting 2019, Conference on the Economics of Crime and Justice 2019, UC Irvine, University of Chicago Economics, University of Chicago Harris School of Public Policy, the 28th Annual Meeting of the American Law and Economic Association, UC Berkeley Labor Seminar, UC Berkeley Public Finance Lunch Seminar, All California Labor Conference 2018, and the 13th Annual Conference on Empirical Legal Studies for helpful comments. We gratefully acknowledge financial support from the Center for Equitable Growth. Yotam Shem-Tov also acknowledges funding from the U.S. Bureau of Justice Statistics.

1 Introduction

Since the 1980s, the United States' incarceration rate has more than tripled. The U.S. now spends $80 billion a year to incarcerate more individuals per capita than any other OECD 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). First, incarceration temporarily "incapacitates" individuals by removing them from society and thus making it more difficult to commit crime. Second, it can influence individuals' criminal behavior post release. Incarceration can rehabilitate (Bhuller et al., 2018a) or deter (Becker, 1968; Drago et al., 2009) convicted individuals, but it can also serve as a "crime school" by exposing them to criminal peers (Bayer et al., 2009; Stevenson, 2017). Moreover, the stigma attached to incarceration might disconnect individuals from the labor market once they are released, further increasing criminal behavior (Grogger, 1995; Kling, 2006; Raphael, 2014; Mueller-Smith, 2015; Agan and Starr, 2018).

This paper studies the causal effect of incarceration on reoffending. A key objective is to separate the incapacitation effects associated with an initial sentence from any behavioral effects of time served on crime committed after release. We begin by estimating the combined effects of incarceration on reoffending through both the incapacitation and behavioral channels. These estimates are a key input to crime control policy decisions. The second part of the study seeks to empirically disentangle the incapacitation and behavioral channels while accounting for both non-linear and heterogeneous effects of exposure to incarceration.

Both analyses require variation in incarceration length that is uncorrelated with individuals' unobserved criminality. To isolate such variation, we use discontinuities in North Carolina's sentencing guidelines, which define permissible punishments according to individuals' offense severity and a numerical criminal history score. Guideline sentences change discretely when scores cross critical thresholds, providing shifts in the sentence type (incarceration vs. probation) and sentence length for otherwise comparable individuals. For example, offenders convicted of first degree burglary face a 30 p.p. jump in the likelihood of incarceration between 4 and 5 criminal history points — a difference that can arise due to quasi-random factors such as whether two prior misdemeanor charges were disposed in the same or consecutive calendar weeks. Although convicted charges are potentially manipulable through plea bargaining, we show that our results are robust to using either the arrested, charged or convicted offense to define the instruments.

We begin with a regression discontinuity (RD) analysis that estimates the effect of incarceration length on reoffending post conviction using two-stage least squares (2SLS). These estimates capture the quantity and type of crime averted by putting offenders behind bars rather than on supervision in the community (i.e., probation) in the years after sentencing. We find that one year of incarceration reduces the likelihood of committing a new offense by 9.5 p.p. (↓22%), a new assault crime by 2.59 p.p. (↓38%), a new property offense by 3.92 p.p. (↓24%), a new drug offense by 3.37 p.p. (↓20%), and being reincarcerated by 16.6 p.p. (↓36%) over the three years after sentencing. This crime reduction persists over longer windows and is still evident even eight years after sentencing.

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 incapacitated that 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. When incapacitation rates are high, monthly offending rates are correspondingly lower. 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. However, due to the initial incapacitation spell, incarceration still causes a reduction in cumulative measures of crime such as *ever* reoffending in the eight years after sentencing.

We also present estimates using measures of reoffending starting at each individual's "at-risk date," which is the date of conviction for those sentenced to probation and the date of release for those sentenced to incarceration. The logic of this approach is that by holding constant the length of time an individual is in the community and at risk to reoffend, any incapacitation effects are removed (Nagin and Snodgrass, 2013; Mears et al., 2016; Harding et al., 2017). We find no evidence of any criminogenic impacts of incarceration. An additional year of exposure reduces reoffending within three years of an individual's at-risk date by either 8.9 p.p. (↓19%) or 0.46 p.p. (↓1%), depending on how reoffending is measured.

While informative, our first set of analyses is limited by several key factors. First, our 2SLS estimates are parameterized by a single endogenous variable (months of incarceration), ruling out any non-linearities in effects. These estimates imply, for example, that shifting an offender from zero to three months has the same impact as lengthening a five-year sentence by an additional three months. Second, treatment effects are likely to be heterogeneous across the individuals we study. 2SLS estimates of models that allow for non-linearity in the effects of incarceration (e.g., a polynomial in length of exposure) require shutting down any such heterogeneity. And third, as we discuss below, 2SLS estimates that measure reoffending from at-risk recover a mixture of the behavioral effects of incarceration and other time-varying factors (e.g., age, aggregate crime rates, etc.). Controlling for these factors is challenging, since they are directly affected by the time of release and are thus endogenous to the treatment.

In the second part of our study, we present a new framework that overcomes these core challenges and allows us to unpack the reduced form evidence. This framework consists of a semi-parametric model of incarceration length and recidivism that describes how the latent propensity to commit crime varies with incarceration exposure, release date, and unobserved criminality. The model allows us to parse the weighted average of effects identified by 2SLS and separately identify both non-linearity and heterogeneity in the effects of incarceration on incapacitation and criminal behavior post release, a key issue when examining possible sentencing reforms. This empirical strategy also allows us to correct for any endogeneity that is induced by measuring reoffending from *at-risk* while directly controlling for time-varying factors such as age and year of release.

We estimate the model parameters via a two-step control function estimator (Heckman and Robb, 1985; Meghir and Palme, 1999; Wooldridge, 2015). First, we estimate a single index ordered-choice model of assignment to different lengths of incarceration by maximum likelihood. The generalized residuals (Gourieroux et al., 1987) are used to proxy for the latent factor that generates omitted variable bias in the assignment to incarceration. Second, we eliminate the initial period of incapacitation by resetting the starting point from which reoffending is measured to each individual's at-risk date. To isolate the behavioral effects of incarceration from any time-varying factors, we model the likelihood of reoffending within t months from release as a function of time-invariant covariates, time-varying factors, a control function, and a flexible function of incarceration length.

The model is then estimated separately for each t using ordinary least squares (OLS). Since the model specifies reoffending rates for any level of incarceration exposure and time horizon, it is straightforward to back out implied incapacitation effects as the change in reoffending from reducing time at-risk while holding incarceration exposure and other factors constant. Although the model makes restrictions on the data generating process, it also has strong testable implications that can be used to validate these assumptions. For example, we show that the model can replicate RD estimates on reoffending within t months from conviction.

Treatment effect heterogeneity is incorporated by interacting the control function with incarceration exposure, allowing effects to vary across individuals with different latent criminal propensities. The resulting model exhibits "essential heterogeneity" in that it links the propensity to participate in a treatment to the treatment effect (Heckman and Vytlacil, 2005, 2007; Heckman and Leamer, eds, 2007). Although we apply our model to the case of incarceration, the structure is broadly applicable to any setting in which treatment involves an initial incapacitation spell and the researcher seeks to estimate effects on behavior afterwards such as job-training programs (Ham and LaLonde, 1996; Eberwein et al., 1997) or military service.

The results show that incarceration has modest crime-reducing behavioral effects. Specifically, one year of incarceration reduces the likelihood of reoffending within three years of release by 7% to 22%, depending on the measure of reoffending. Using the model estimates to decompose our quasi-experimental estimates of the effects on reoffending measured from conviction, we find that within one year of conviction, the majority of the reduced form effects can be explained by incapacitation alone, with the behavioral channel explaining between 0.05% and 10% of the total reduction in ever reoffending. Within five years of conviction, however, the importance of the two channels is reversed, with behavioral effects now explaining between 30% and 84% of the reduction in reoffending. Our estimates also show that the majority of the behavioral impacts come from the first year of incarceration, with limited effects of additional exposure beyond that. An analysis of treatment effect heterogeneity also finds evidence of selection on gains: the behavioral effects are largest for highest risk offenders (i.e., those currently sentenced to the longest incarceration spells).

Our estimates provide critical inputs for optimal crime control policy decisions. To summarize the implications, we conduct a simple cost-benefit analysis that provides a useful benchmark for the value of the estimated crime reduction relative to the cost of incarceration. On average, an additional month of incarceration reduces cumulative reoffending after eight years by -0.0298 new offenses, while costing roughly \$2,738 in correctional spending. To break even, the marginal averted offense would therefore need to be valued by society at roughly \$92,000 ($= \frac{\$2{,}738}{0.0298}$). For felony offenses, the break-even value is roughly \$164,000. An alternative cost-benefit analysis that assigns a dollar value to each reoffending event suggests that the cost of incarceration is higher than the value of the crime averted. Thus, despite a large estimated reduction in crime from more aggressive sentencing through both incapacitation and behavioral channels, the high cost of incarceration likely outweighs the social benefit of lower crime.

We contribute to a broad literature across the social sciences on the relationship between incarceration and reoffending.[1] In recent years, a common approach to the problem of selection to incarceration based on latent criminality has been to take advantage of random or rotational assignment of defendants to judges.[2] A few papers utilizing this design are closely related to our study. Bhuller et al. (2018a) find that prison sentences have substantial rehabilitative effects among Norwegian criminal defendants. Their approach to separating incapacitation from rehabilitative effects is to examine new offenses 25 months after initial conviction and beyond, when the initial sentence no longer influences incarceration status. We provide the analogous estimate in our context, which shows a zero effect. Mueller-Smith (2015), meanwhile, finds large criminogenic effects of incarceration length on the likelihood of offending among criminal defendants in Harris County, TX. Mueller-Smith uses a panel regression model with multiple endogenous variables for current incarceration status, release from incarceration, and a cumulative measure of incarceration exposure. These results show moderate incapacitation effects and large criminogenic effects of incarceration, generating net increases in the frequency and severity of recidivism.

Our estimates are similar in sign but smaller in magnitude than Bhuller et al. (2018a) and differ in both sign and magnitude from Mueller-Smith (2015). This may reflect differences in the causal

[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); Raphael and Lofstrom (2015). Miles and Ludwig (2007) provides a review of the evidence from the Criminology literature on incapacitation effects.

[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), Mueller-Smith (2015), Aizer and Doyle (2015), Stevenson (2016), Harding et al. (2017), Zapryanova (2017), Arnold et al. (2018), Arteaga (2018), Aneja and Avenancio-León (2018), Bhuller et al. (2018a), Bhuller et al. (2018b), Dobbie et al. (2018b), Dobbie et al. (2018a), Huttunen et al. (2019), Norris (2018), and Norris et al. (2018).

effects of prisons in Norway, Harris County, and North Carolina or differences in the treatment effects for those shifted to incarceration in each experiment. The latter type of treatment effect heterogeneity can lead to substantial variation in estimates across research designs. 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.

We build upon and extend both Bhuller et al. (2018a) and Mueller-Smith (2015) in several ways. The multiple discontinuities we exploit provide variation in both the extensive and intensive margin effects of incarceration, allowing us to estimate non-linear impacts of incarceration on reoffending. In addition, our semi-parametric selection model provides a new framework for separately identifying the incapacitation and behavioral channels under treatment effect heterogeneity.

Papers exploiting non-judge variation also find contrasting effects. Kuziemko (2013), for example, compares a parole system to a fixed-sentence regime and argues that each additional month in prison reduces three-year reincarceration rates by 1.3 p.p. for a sample of parolees in the state of Georgia. On the other hand, Franco et al. (2017) find that reincarceration rates are higher for initially incarcerated offenders. Differences in the institutional setting and the impact of accounting for technical revocations of probation and parole can potentially explain some of these differences. We discuss this issue in detail below and propose possible solutions.[3]

A final strand of related literature uses exogenous shocks to prison populations to identify the relationship between incarceration rates and crime.[4] This type of variation captures effects that go beyond the partial equilibrium analysis we study in this paper. Nevertheless, estimates from this literature also vary widely (Levitt, 1996; Raphael and Lofstrom, 2015).

The remainder of this paper is organized as follows. Section 2 presents the conceptual framework, discusses common empirical approaches for measuring reoffending, and describes the interpretation of instrumental variable (IV) estimates in this context. Section 3 describes the institutional setting and the data used. Section 4 describes the empirical strategy for identifying causal effects. Section 5 presents results from the IV analysis. Section 6 discusses possible threats to identification and how we overcome them. Section 7 lays out our empirical strategy for separating incapacitation and behavioral effects and reports the results of this approach. Section 8 discusses the policy implications of both analyses. Section 9 concludes.

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

[4] Notable example include Marvell and Moody (1994); Levitt (1996); Drago et al. (2009); Maurin and Ouss (2009); McCrary and Sanga (2012); Buonanno and Raphael (2013); Barbarino and Mastrobuoni (2014) and Raphael and Lofstrom (2015).

2 Conceptual framework

We begin by formalizing the causal parameters of interest in the language of potential outcomes. This serves both to clarify the estimates presented below and to illuminate some of the unique identification challenges faced in this context. We present a simplified model throughout, suppressing all covariates X_i and examining the case of a single binary instrument Z_i and a discrete ordered treatment $D_i \in \{0, 1, \ldots, \bar{D}\}$ (i.e., incarceration length) with potentially both non-linear (Løken et al., 2012; Lochner and Moretti, 2015) and heterogeneous effects.

2.1 Outcome measurement

Suppose we observe a panel of offenders for T periods after conviction for an initial offense. In each period we observe if the offender was arrested and whether the offender is incapacitated in prison or jail at period t. Since estimates of effects on offending at time t (e.g., in a given month) can be imprecise, the literature has focused on estimating the effect of incarceration exposure on the length of time until an offender commits a new offense. These estimates reflect effects on "failure functions" in the terminology of duration analysis. A key question in analysis of reoffending failure times is when to begin measurement. Two starting points have been used in the literature: (i) the date of conviction; and (ii) the date at which the offender is released back to the community and is therefore at-risk to reoffend.

To capture these two distinct approaches, we model failure functions as potential outcomes. Specifically, let $Y_{i,t}(d)$ be an indicator for whether individual i would reoffend within t months of his initial conviction date if initially incarcerated for d months. Since this object measures time to reoffend from the conviction date—when incarceration sentences are assigned—the amount of time an individual is at-risk to reoffend is given by $t - d$. An indicator for reoffending within t months from the date of release can therefore be expressed using a shift of d months in the reoffending window: $Y_{i,t+d}(d)$.

2.2 Identification using an IV

The IV approach accounts for the fact that incarceration assignment is unlikely to be independent of individuals' propensity to reoffend. In our setting, the instruments are indicators for being above a discontinuity in punishment (described in full detail below), which we denote $Z_i \in \{0, 1\}$. Let $D_i(z)$ denote the number of months individual i would be incarcerated when she is to the left ($z = 0$) or to the right ($z = 1$) of a punishment discontinuity. Realized incarceration sentences are $D_i = D_i(1)Z_i + D_i(0)(1 - Z_i)$ and observed reoffending can be written as $Y_{i,t} = Y_{i,t}(D_i(1))Z_i + Y_{i,t}(D_i(0))(1 - Z_i)$.

The instruments are assumed to satisfy the usual assumptions of the LATE framework (Imbens and Angrist, 1994; Angrist et al., 1996) and its extension to treatments with multiple levels (Angrist and Imbens, 1995)

Assumption 1. (First stage) $\mathbb{E}[D_i|Z_i = 1] > \mathbb{E}[D_i|Z_i = 0]$

Assumption 2. (Exogeneity) $Y_{i,t}(d), D_i(1), D_i(0) \perp\!\!\!\perp Z_i \quad \forall\, d \in \{0, 1, 2, \ldots, \bar{D}\}$

Assumption 3. (Monotonicity) $D_i(1) \geq D_i(0) \quad \forall i$

Assumption 1 states that the instrument influences exposure to treatment. In our setting, this assumption says that individuals to the right of a discontinuity face a harsher sentencing regime. Assumption 2 implies that the instrument is orthogonal to individuals' latent criminal propensities. That is, conditional on the running variable, individuals to the left and to the right of a discontinuity are comparable. Assumption 3 states that being above a punishment discontinuity weakly increases the sentences of all offenders, an assumption that is highly plausible given the role of guidelines in the sentencing process. We test one implication of the monotonicity assumption in a setting with a multi-valued treatment in Section 4.1.

A complier in our setting is an individual that is incarcerated for longer because she is located to the right of a punishment discontinuity, i.e., $D_i(1) > D_i(0)$. There are many types of compliers corresponding to different shifts in exposure to incarceration. We define a type d complier as an individual shifted to at least d months of incarceration due to the instrument, i.e., $D_i(1) \geq d > D_i(0)$.

In the existing literature, the impacts of incarceration on future outcomes are usually modeled using either the length of incarceration D_i or a binary indicator for any incarceration, $\mathbb{1}\{D_i > 0\}$. When using D_i as the endogenous variable, IV recovers the "average causal response" discussed in Angrist and Imbens (1995):

$$\beta_{\text{conviction}}(t) \equiv \frac{\mathbb{E}[Y_{i,t}|Z_i = 1] - \mathbb{E}[Y_{i,t}|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}\left[Y_{i,t}(d) - Y_{i,t}(d-1) | \underbrace{D_i(1) \geq d > D_i(0)}_{\text{Type } d \text{ compliers}}\right] \quad (1)$$

where

$$\omega_d = \frac{\Pr(D_i(1) \geq d > D_i(0))}{\sum_{l=1}^{\bar{D}} \Pr(D_i(1) \geq l > D_i(0))} \quad (2)$$

Equation (1) shows that the IV estimand $\beta_{\text{conviction}}(t)$ is a weighted average of causal effects for different populations of compliers. For example, $\mathbb{E}[Y_{i,t}(d) - Y_{i,t}(d-1) \mid D_i(1) \geq d > D_i(0)]$ is the effect of an additional month of incarceration on the likelihood of reoffending within t months of conviction for individuals that would be incarcerated for strictly less than d months when $Z_i = 0$, but otherwise would be incarcerated for at least d months.[5] Although $\beta_{\text{conviction}}(t)$ recovers a

[5]In Appendix B, we discuss how IV estimates with respect to the the failure function $Y_{i,t}$ can be represented as a summation of the effects on the hazards of reoffending at period t conditional on not reoffending prior to time t.

combination of incapacitation and behavioral effects, this composite impact is the policy-relevant parameter when evaluating the overall effect of incarceration length on reoffending in the years after conviction.

In an attempt to separate incapacitation effects from any behavioral effects of incarceration after release, it is common to use the same IV procedure but with reoffending measured from the date of release. The estimand that is now recovered by IV, without adjusting for any time-varying controls, is

$$\begin{aligned}\beta_{\text{at-risk}}(t) &\equiv \frac{\mathbb{E}\left[Y_{i,t+D_i}|Z_i=1\right]-\mathbb{E}\left[Y_{i,t+D_i}|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]} \\ &= \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}\left[Y_{i,t+d}(d)-Y_{i,t+d-1}(d-1)|D_i(1)\geq d > D_i(0)\right]\end{aligned} \tag{3}$$

A sketch of the proof of Equation (3) is presented in Appendix C. The estimand in Equation (3) captures a mixture of the behavioral effects of incarceration and the effects of other time-varying factors, since incarcerating offenders for an additional year exposes them to prison for a longer period but also makes them older by an additional year and releases them into a different environment, which may also influence reoffending

Without imposing additional structure, it is not possible to express treatment effect estimates as the sum of a behavioral effect and a bias term due to aging (or other time varying factors). In fact, there is no non-parametric potential outcomes representation of the behavioral effect because it does not correspond to a well-defined hypothetical manipulation—you cannot incarcerate someone for an extra year without also making them older.[6] Thus, in at-risk estimates, any time variation in individual characteristics (e.g., age) or the environment (e.g., overall crime rates) will be fully attributed to the effects of incarceration length. Importantly, simply controlling for time-varying factors such as age at the date of release or local unemployment rates at release can potentially lead to bias since these variables are functions of incarceration length and are therefore also endogenous.[7]

Finally, researchers often use a binary indicator for whether the individual was incarcerated or not, $\mathbb{1}\{D_i > 0\}$, as the endogenous treatment. Angrist and Imbens (1995) showed that when the treatment is multi-valued, this specifications yields biased estimates of $\beta_{\text{conviction}}(t)$. In Appendix B, we show that this estimand can also be interpreted as a linear combination of the *extensive* and *intensive* margin impacts of incarceration on an outcome of interest. Extensive effects are those for individuals who counterfactually would have received no incarceration sentence ($D_i(1) > D_i(0) = 0$). Similarly, intensive margin effects reflect lengthening incarceration for individuals who

[6] In Appendix C, we present an example for an explicit model of potential outcomes, which illustrates how the estimand reflects a mixture behavioral responses and time-varying factors.

[7] In Appendix C, we present an example of how controlling for such time-varying factors leads IV estimators to identify an estimand without a clear causal interpretation due to conditioning on an endogenous variable, which leads to a type of post-treatment adjustment bias (Rosenbaum, 1984). This issue is one of the motivations for the selection model based analysis in Section 7.

otherwise would have spent less (but not zero) time behind bars ($D_i(1) > D_i(0) > 0$). However, if the instrument has no intensive margin effects (i.e., $\Pr(D_i(1) > D_i(0) > 0) = 0$), then using a binary indicator as the endogenous variable admits a clear causal interpretation: it identifies the average effect of any incarceration sentence for individuals shifted to incarceration due to the instrument.[8]

Our IV analysis below proceeds in two parts. We begin by describing the reduced form effects of the discontinuities on reoffending, i.e., the relationship between $Y_{i,t}$ and Z_i. Second, we report 2SLS estimates of $\beta_{\text{conviction}}(t)$ and $\beta_{\text{at-risk}}(t)$ that scale the reduced form effects by the shifts in incarceration length caused by the instrument. We do not present 2SLS estimates using any incarceration as the endogenous variable, since, as we show below, our instruments shift exposure to incarceration through both the extensive and intensive margins.

2.3 Technical probation violations and competing risks

Virtually all individuals *not* sentenced to incarceration are instead given a probation term that restricts alcohol and drug use, work and socializing, and travel and requires the payment of court fees and fines. Individuals who violate the terms of their supervision can be incarcerated as a result. These probation "revocations" are frequently not associated with any new crimes, making it unclear whether to classify such instances as reoffending. However, probation officers may also revoke individuals they suspect are involved in new criminal activity. For example, Austin and Lawson (1998) found 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 (Kuziemko, 2013; Yang, 2017), although many studies do not discuss the issue explicitly.

Reincarceration due to technical revocations can bias incarceration effects estimates in two ways. First, if revocations mask genuine criminal activity, not counting them as reoffending may artificially deflate reoffending rates in the probation (i.e., untreated) population. Second, even if technical revocations are not associated with actual crimes, revoked individuals may have otherwise committed crimes in the future. Since these individuals go to prison, overall offending in the control population will go down. If those revoked are also higher risk on average, the remaining control units at-risk to reoffend may be positively selected, exacerbating the problem.[9]

To account for these issues, we adopt four approaches. First and foremost, we present estimates of the effects of incarceration on reoffending with and without including probation revocations. Second, in Appendix E, we report estimates assuming that the risks of probation revocations and

[8]Note that the null $H_0 : \Pr(D_i(1) > D_i(0) > 0) = 0$ can be empirically examined by testing the following necessary condition that most hold if the null is true: $H_0 :\ \Pr(D_i(1) \geq 1 > D_i(0)) \geq \Pr(D_i(1) \geq d > D_i(0))\ \forall d > 1$. This is a necessary condition, and not a sufficient condition, for the null to be satisfied.

[9]In a recent literature review and replication analysis, Roodman (2017) discusses how technical parole violations can impact the estimated effects of incarceration length on reincarceration from Kuziemko (2013) and Ganong (2012). Roodman refers to such impacts as "parole bias".

committing new offenses are independent. Under this assumption, we can simply drop from the analysis any observations in which a technical revocation occurred before a new offense and before period t.[10] In practice, we view these independent risk estimates as an upper bound, since it seems unlikely that probation revocations are *negatively* correlated with risk, i.e., that the least dangerous individuals are being revoked. Third, we also derive non-parametric bounds analogous to those in the competing risks literature (Peterson, 1976). This approach is laid out in Appendices E and F. And fourth, we drop individuals censored by probation revocations but use a control function approach to correct for any resulting sample selection (Section 7 and Appendix M.5).

3 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.

3.1 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 (hereinafter SSA). These guidelines were crafted as part of a nationwide shift towards 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 (U.S. Department of Justice, 1996). By 2008, the number of states with sentencing guidelines had increased to 28 (National Center for State Courts, 2008). Sentencing guidelines have been used elsewhere to estimate effects of features of the criminal justice system.[11]

The SSA eliminated parole by requiring that defendants serve the entirety of a minimum sentence.[12] 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.[13] Felony offenses are grouped into ten different classes based on severity of the offense. Offenders are assigned a criminal history score (referred to as "prior record points") that assigns 1 point for some

[10] This is because $\mathbb{E}\left[Y_{i,t}(d)|R_{i,t}\right] = \mathbb{E}\left[Y_{i,t}(d)\right]$, where $R_{i,t}$ denotes an indicator for whether individual i had a probation revocation prior to committing a new offense up to date t.

[11] Related designs have been studied by Kuziemko (2013) and Ganong (2012) for the case of parole, Hjalmarsson (2009) for juvenile offenders, and Chen and Shapiro (2007) for the case of 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.

[12] After doing so, defendants become eligible for early release, but can serve no more than 120% of their minimum sentence. Most offenders are released close to their minimum.

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

misdemeanor offenses and 2-10 points for previous felony offenses, depending on the seriousness of the offense. When an individual was previously convicted of multiple offenses in the same *calendar* week, only the most serious offense is used. Additional points are added if offenses are committed while the offender is on supervision or all the "elements" of the current offense are included in any prior offenses. As a result, two individuals with highly similar criminal histories can have different prior record scores depending on the timing and precise nature of their previous offenses.

The SSA groups individuals into prior record "levels" according to their total points and sets minimum sentences for each offense class and prior record level combination, which we refer to as a grid "cell."[14] This is visually illustrated in Figure 1, which shows North Carolina's official grid with annotations. Each grid cell is assigned 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;[15] and (iii) community punishment, or regular probation. These sentence types are denoted with "C/I/A" lettering at the top of each cell in the grid. For more details, see the official sentencing guidelines for the years 1994 to 2013 in Appendix D.

The combination of shifts in required sentence lengths and allowable sentence types generates large differences in punishments meted out across the grid as is shown in Figure 1 (and Appendix D). For example, offenders with 9 prior points and a Class I charge can be given an incarceration sentence, whereas offenders with 8 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 placement to validate our research design.

3.2 Data sources

We use administrative information on arrests, charges, and sentencing from two sources. The first is records provided by the North Carolina Administrative Office of the Courts (AOC) covering 1990 to 2017. These data includes rich information on defendants, offenses, convictions, and sentences for all cases disposed in Superior Court, which hears felony offenses. This data is used to measure the set of initial charges associated with a conviction and to construct reoffending measures.

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, sanctions imposed, and incarceration spells in jails and prisons. The data also contain reliable

[14] 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.

[15] Intermediate can also include "shock" probation, which includes a short incarceration spell before probation begins.

measures of probation revocation and additional details on offenders' demographics, including age, height, weight, languages spoken, race, and ethnicity. We use this data to construct our instruments and to measure incarceration.

Our primary measure of reoffending is constructed using both AOC and DPS records and counts the number and type of new criminal charges (or convictions) filed against an individual in Superior Court at a given time period. 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. We date new charges (or convictions) using the date of offense, rather than the date charges were filed, in order to eliminate any delays due to lags in detection in our court proceedings. We also consider alternative measures of reoffending such as only any new convictions recorded in either the AOC or DPS data, the type of new criminal charges (e.g., assault, property, drug) or whether the defendant was returned to incarceration for either a new offense or a probation revocation.

3.3 Sample construction and restrictions

Offenders routinely face multiple charges simultaneously and can be sentenced to concurrent incarceration spells for offenses that were committed at different dates. To overcome this issue, we conduct the 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.[16]

Because our research design utilizes discontinuities in felony sentencing guidelines, the analysis sample is restricted to individuals convicted for felony offenses committed between 1995 to 2014 and therefore sentenced on the felony grid. We do not include misdemeanors or DWIs, since they are sentenced under different guidelines. We drop observations in which the individual is incarcerated at the time of sentencing due to a probation revocation or a concurrent charge, since these sentences are unlikely to be affected by our instruments.

We focus on Class E through Class I offenses (92.3% of the observations) and include individuals with prior record points of 25 or fewer, which captures the vast majority of offenses. This restriction is motivated by the fact that in each of these five classes of offense severity there are discontinuities both in the type and length of punishment, as is discussed in Section 4.1. However, when using classes more severe than E (e.g., Classes D and C), there are only discontinuities in the guidelines with respect to the intensive margin and no discontinuities in the extensive margin of

[16] Another approach would be to group charges into "cases" where either the conviction, offense, or 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.

the punishment type.[17] Finally, we also restrict the analysis to individuals aged between 15 and 65 at the time of offense.

3.4 Descriptive statistics

Summary statistics for our sample are presented in Table 1. On average, offenders are predominately male, roughly 50% black, and 30 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 4.4 months. Conditional on receiving an incarceration sentence the average length is 13.6 months. As a state North Carolina has similar incarceration and recidivism rates as the overall rates in the U.S. (see Figure 1 in Norris et al. (2018)).[18]

Roughly 55% of the sample reoffends at some point in the period we study. Most offenders who reoffend do so in the first few years after being released. 48% of offenders reoffend within five years of release, and 33% reoffend in the first 2 years. Appendix Figure A.1, Panel (a), shows the likelihood of committing any new offense up to time t. The slope of this function is the hazard rate. Panel (b) shows the hazard rate, which is the likelihood of committing an offense at time t conditional on not reoffending prior to that time. The hazard function decreases sharply over time, indicating that as time goes by the likelihood that an offender who did not reoffend will commit a crime is decreasing. A key objective of our study is to understand how the failure function is influenced by exposure to incarceration.

Appendix Figure A.2 compares the failure functions of reoffending for initially incarcerated (blue line) and non-incarcerated (red line) individuals. Panel (a) shows that when technical probation revocations are *not* counted as reoffending the individuals who where sentenced to incarceration have higher rates of criminal involvement once released. However, if such technical violations are considered as a signal of crime, then this difference reverses. This descriptive pattern demonstrates the importance of reporting estimates of incarceration effects on reoffending both with and without including probation revocations.

4 Empirical strategy

Our research design exploits the extreme non-linearities in sentencing outcomes when moving horizontally across the boundaries of SSA 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 to exploit. Each discontinuity has 4 to 5 unique values of the running variable (prior points) on either side except the first one, which has only one prior point value to its left and which consequently we do not

[17] Including Class D and C in the analysis does not alter any of our results.

[18] The demographic characteristics of our sample population are similar to the ones observed in other studies. Appendix Table A.1 shows the average (or median) age of offenders across other studies, which is usually around 30.

use as an excluded instrument. Our setting is thus not a classic RD scenario with a continuous running variable like a congressional election (Lee, 2008) or a college loan program (Solis, 2017). Instead, we have a discrete running variable; our specification therefore reflects a parametrized RD design (Clark and Del Bono, 2016).[19]

Specifically, our model includes separate linear slopes in each cell of the sentencing grid and allows for vertical shifts—or "jumps"—between horizontally adjacent cells. Figure 2 Panel (a) visually illustrates this idea for one offense class (F) 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 (when prison punishments become allowed).

Our preferred estimator "stacks" all the variation at cell boundaries in each offense class to estimate a single treatment effect. This estimator is written formally in the two-equation system below. Equation (4) (first stage) estimates length of incarceration D_i as a function of prior points, convicted charge severity, punishment discontinuities, and other covariates. Equation (5) represents the relationship between reoffending within t months from conviction $Y_{i,t}$, incarceration and grid (and offender) controls. The system is estimated using 2SLS.

$$D_i = \underbrace{\eta^1_{c_i} + X'_{it}\alpha_1}_{\text{Baseline controls}} + \underbrace{\sum_{k\in classes} 1\{class_i = k\}\left[\sum_{l\in thresh} \beta^2_{lk} 1\{p_i \geq l\}(p_i - l + 0.5) + \beta^1_k p_i\right]}_{\text{Linear slopes in prior points by class and level}} \tag{4}$$

$$+ \underbrace{\sum_{k\in classes}\sum_{l\in thresh\neq 0} \gamma^2_{kl} 1\{p_i \geq l\} 1\{class_i = k\}}_{\text{Prior record level discontinuities}} + \underbrace{\sum_{k\in classes} \gamma^3_k 1\{p_i \geq thresh_0\} 1\{class_i = k\}}_{\text{Absorb level 0 discontinuity}} + \epsilon_i$$

$$Y_{i,t} = \beta_0 D_i + \underbrace{\eta^1_{c_i} + X'_{it}\alpha_1}_{\text{Baseline controls}} + \underbrace{\sum_{k\in classes} 1\{class_i = k\}\left[\sum_{l\in thresh} \beta^2_{lk} 1\{p_i \geq l\}(p_i - l + 0.5) + \beta^1_k p_i\right]}_{\text{Linear slopes in prior points by class and level}} \tag{5}$$

$$+ \underbrace{\sum_{k\in classes} \gamma^3_k 1\{p_i \geq thresh_0\} 1\{class_i = k\}}_{\text{Absorb level 0 discontinuity}} + e_{i,t}$$

where D_i is the length of incarceration that the offender served, $\eta^1_{c_i}$ and $\eta^2_{c_i}$ are offense class (e.g., E, I, G) specific fixed effects, p_i is prior points, and X_i is a vector of control variables. The

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

thresholds refer to the prior record boundary levels in place at the time of the offense (e.g., 5 or 9 points), with $thresh_0$ denoting the first boundary (i.e., 1 or 2 points), which we do not use as an instrument. When estimating the changes in slope on either side of each boundary (the $1\{p_i \geq l\}(p_i - l + 0.5)$ effects), 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] X_{it} includes 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.

The specification above uses the indicators for being to the right of each discontinuity (that is, the $\gamma_{kl} 1\{c_i = k\} 1\{p_i \geq l\}, \forall l \in [5, 9, 15, 19], \forall k \in classes$) as the instruments. Among the 20 instruments at our disposal, five correspond to parts of the grid where the punishment type varies (e.g., when an incarceration sentence is first allowed) as is illustrated in Figure 1. For the main analysis, we use these five punishment type discontinuities, which provide the most salient changes in sentences (red lines in Figure 1). When exploring heterogeneity in treatment effects, we also use the other 15 discontinuities to maximize variation. We demonstrate, however, that results are similar regardless of the instrument set used, including if the five punishment type discontinuities are not used at all.

4.1 First stage effects of cell discontinuities

This research design captures large discontinuities in sanctions across the sentencing grid. For example, Figure 2 Panel (a) shows that an offender convicted of a class F felony offense (which includes assault with serious bodily injury) faces a 34 p.p. increase in the probability of incarceration between 8 and 9 prior points, which determines whether the offender is classified to prior record level III or IV. Appendix Figure A.3 examines the other offense classes and documents multiple discontinuities in the type and length of punishment. Note that this variation occurs at different values of prior record points depending on the offense class. For example, in class H, which contains the most defendants in the data, the change in punishment type falls between prior record levels V and VI, which generates an extensive margin discontinuity between prior record points 18 and 19. This discontinuity falls at much lower values of prior points in class F.

Another way to visualize the shifts in time served due to our instruments is to plot the weights of the average causal response from Equation (2) ($\Pr(D_i(1) \geq d > D_i(0))$). These weights capture the distributional shifts in exposure to incarceration caused by each of the binary instruments and can be estimated as $\mathbb{E}[1(D_i \geq d)|Z_i = 1] - \mathbb{E}[1(D_i \geq d)|Z_i = 0]$. Figure 2 Panel (b) plots the $\widehat{\Pr}(D_i(1) \geq d > D_i(0))$ estimates for class F and shows that being above the discontinuity generates a shift in the entire distribution of incarceration exposure. Appendix Figure A.4 plots

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

the $\widehat{\Pr}(D_i(1) \geq d > D_i(0))$ for the other felony classes. The instruments provide wide variation in exposure to incarceration. Estimates of $\Pr(D_i(1) \geq d > D_i(0))$ also provide a test for the monotonicity assumption (Angrist and Imbens, 1995). If the instruments satisfy monotonicity then $\widehat{\Pr}(D_i(1) \geq d > D_i(0))$ should never cross the x-axis at zero, since a probability cannot have a negative value. Appendix Figure A.4 confirms that all the instruments pass this validity check.

In the regressions that follow, we control in a flexible way for the offender's criminal history using both the linear controls in prior points from the RD specification as well as indicators for any previous incarceration spell (included in the X_{it}), 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 still provides strong variation in the type and length of punishment, as shown by the first stage F-statistics presented below each of the results tables. The instrumental variation therefore primarily comes from the non-linear mapping between prior convictions and prior record points, as opposed to simple counts of prior convictions.

4.2 Instrument validity

As is standard in instrumental variable designs, it is important that the instruments are uncorrelated with unobserved confounders. In our setting, it is critical that individuals' latent criminality evolves smoothly across each discontinuity. In this section, we perform a series of balance and validation exercises demonstrating that our instruments do not predict individual characteristics, supporting the assumption that conditional on our controls for prior points individuals just to the left and just to the right of each discontinuity provide valid treatment and control groups to assess the causal effects of incarceration. Since there are many relevant pre-treatment covariates, we make use of a predicted reoffending (risk) score calculated by regressing an indicator for reoffending on all the pre-treatment covariates (using only non-incarcerated offenders) and fitting predicted values for all offenders.[21]

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 Wald test for the joint significance of all five discontinuities also fails to reject zero effects (the p-value is 0.159, with an F-statistic of 1.58 and 5 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.

Several pieces of evidence 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

[21]Summarizing imbalance by the covariates' relationship to the outcome surface is a common methodology in the literature (Bowers and Hansen, 2009; Card et al., 2015; Londono-Velez et al., 2018). We also experimented with using more sophisticated (i.e., machine learning models) to construct the risk score; the results are similar.

means between the points straddling the discontinuities (see Appendix Figure G.1). In other words, although sentences change abruptly across consecutive prior points at the discontinuities in punishment type, (Appendix Figure G.2), the covariates do not.

Second, a 2009 reform to the grid shifted each discontinuity one prior point to the left or right. This change shifted the first stage as well, as shown in Appendix Figure A.5. Despite this shift, the distribution of covariates across prior points remained the same. We demonstrate this by estimating Equation (4) in the two years before and after the change, but define the location of each discontinuity using the *old* grid. We then interact the 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 Appendix 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.[22] Large changes in the first stage, therefore, do not lead to changes in covariates, as would be expected if systematic sorting were a concern.

Finally, Appendix Figure G.4 shows that there is no evidence of discontinuities in the density of offenders around punishment type discontinuities. Appendix Figure G.5 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 (the first stage).

Thus, there is no evidence of offenders sorting to avoid harsher punishments and, overall, there is strong support for the validity of our instruments. Nevertheless, after estimating our core results, we conduct additional validity and 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. These tests include demonstrating that our 2SLS estimates are highly robust to the inclusion of a large set of individual controls, reporting estimates using subsets of the instruments, and defining our instruments using charges at arrest.

5 Causal effects of incarceration

In this section, we present results for the effects of incarceration on reoffending in the years after sentencing. These estimates capture a combination of both incapacitation and behavioral effects, are non-parametrically identified, and recover parameters that are of key interest for crime control policy decisions.

[22] Appendix Figure G.3 demonstrates this visually by plotting the distribution of predicted risk scores under the old and new grid.

5.1 Reduced form estimates

We begin with a visual summary of our reduced form evidence by focusing on felony class F and estimating the effects of being to the right of the punishment type discontinuity on various outcomes. Figure 4 shows that individuals to the right of the discontinuity experience a 11.9 p.p. drop in their likelihood of being reincarcerated within three years of conviction. At least part of this decline reflects the fact that individuals to the right of the discontinuity are incapacitated for a large portion of this three year period. To investigate this channel, Figure 5 plots the likelihood of spending any time behind bars *in* a given month since conviction. Individuals to the right of the discontinuity have a sharp increase in the likelihood of being incarcerated at time 0 of 31 p.p., which is exactly the first stage effect (upper-left plot) of our instrument.

Offenders to the right of the discontinuity are also more likely to be incarcerated 6 and 12 months after conviction. However, over time the reduced form difference diminishes and, after 24 months, a difference of less than 3 p.p. remains (bottom-right plot). Figure 5 also shows that the discontinuities stop predicting incarceration status primarily as a result of initially incarcerated offenders being released, causing the dots to the right of the discontinuity to drop down.[23]

Next we examine the dynamic effects of incarceration on reoffending and incapacitation across offenders from all felony classes. To show a single reduced-form effect, we estimate Equation (4) while imposing that the coefficients on the indicators for being to the right of a punishment type discontinuity are all equal (i.e., $\gamma^2_{E,4} = \gamma^2_{F,9} = \gamma^2_{G,14} = \gamma^2_{H,19} = \gamma^2_{I,9} = \gamma^{RF}$). This strategy averages effects across all five offense classes in our analysis dataset, but collapses our variation into a single coefficient.[24]

Figure 6 combines offending and incapacitation outcomes into a single graph that examines effects at each month over the eight years after conviction. Each point in Panel (a) represents an estimate of γ^{RF} for outcomes measured *within* a single month from conviction.

The discontinuities cause a large and immediate increase in incarceration status, 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 30 months, the effect is no longer statistically distinguishable from zero. And after 36 months, the estimates suggest no difference in incarceration rates.

The reduced form effects on committing a new offense and committing a new offense or a probation revocation *within* month t are shown in the red and maroon lines, respectively. There is a negative effect on the probability of reoffending that lasts at least three years after conviction and does not seem to increase afterwards. The fact that the differences in offending stabilize at

[23]The reduced form patterns documented in Figures 4 and 5 are similar across the other felony classes as is shown in Appendix Figures H.1, H.2, H.3, H.4, and H.5.

[24]An alternative approach is to use the average of the five discontinuities $\frac{\gamma^2_{E,4}+\gamma^2_{F,9}+\gamma^2_{G,14}+\gamma^2_{H,19}+\gamma^2_{I,9}}{5}$, which yields highly similar results.

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 red (and maroon) line would lie above zero.

Since period-by-period comparisons are noisily estimated relative to cumulative measures such as committing any new offense, we next examine the reduced form effects on *any* reoffending within t months from conviction in Panel (b) of Figure 6. This graph shows that there is a permanent decrease in the probability of reoffending when measured as committing a new offense and an even larger impact when including probation revocations as reoffending. The difference between these two effects exactly captures the impact of technical probation revocations that occur without any new criminal offenses recorded in the AOC or DPS data.

The decrease reaches a nadir after roughly 18 months, when the estimate begins to increase and continues to do so until 8 years post conviction. After the fifth year the differences seem to stabilize. 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. Many individuals not initially incarcerated, however, have already reoffended, generating the slight increase after the 18 months.

The fact that the red line stabilizes below zero, especially when including probation revocations in our measure of reoffending (the maroon line), is again indicative that an initial term of incarceration does not increase criminal behavior in the long run. The effects on cumulative new offenses show a similar pattern, but the effects stabilize earlier, after roughly three years, as shown in Figure 6 Panel (c).

These estimates recover policy relevant parameters for optimal crime control policy. They estimate the quantity of reoffending averted by incarceration over a period of t months, which is a key input in a cost-benefit analysis of incarceration effects. The estimates are also by no means purely mechanical. It is entirely plausible that incarcerated individuals would have committed relatively little crime if they had been put on probation instead, which is the relevant counterfactual.

Nevertheless, these estimates capture both the impact of incapacitation and behavioral effects on offenders after release. In a first attempt to separate these two channels, Appendix Figure A.6 plots the effects on the cumulative number of new offenses that occurred between 36 and t months from conviction. After 36 months, as shown in Figure 6, the discontinuities no longer predict incarceration status in a given month. Any effects measured after month 36, therefore, cannot be attributed to mean differences in incapacitation. These estimates are relatively precise zeros, suggesting that incarceration does not have any criminogenic effects on reoffending between three and eight years after conviction. Including probation revocations in the reoffending measure has no impact on the estimates. This indicates that the differential impact of probation revocations is in the first three years after conviction.[25]

[25] The approach of estimating incarceration effects on reoffending using a measure of crime that includes only

5.2 2SLS estimates

We next present 2SLS estimates using months of incarceration as the endogenous regressor of interest. Table 2 contains results for committing any new offense within 3 years of conviction. Column 1 shows that the OLS estimate is negative and suggests that one year of incarceration reduces reoffending by 12.9% over that period. Adding controls (Column 2) decreases the coefficient somewhat, which reflects the fact that those assigned incarceration typically have higher recidivism risk according to their baseline covariates. The 2SLS estimates are substantially more negative (over 50%) than OLS, however, suggesting that individuals sentenced to incarceration are negatively selected along unobservable dimensions as well. These estimates find that incarcerating an offender for one year reduces the likelihood of committing any new offense by 22.4% within three years from conviction. Reassuringly, the 2SLS estimates are also stable to the inclusion of flexible controls for criminal history and demographics, which do not have a substantial impact on the estimated effects (Columns 3 and 4).

To investigate the effects on different types of reoffending, in Table 3 we report 2SLS estimates with indicators for committing different types of offenses within three years of conviction as the outcomes. The effects of one year of incarceration are similar in sign and magnitude across offense types relative to their means. For example, a one year incarceration spell reduces the likelihood of committing a new drug offense by -3.372 p.p. (↓20.4%), a new property offense by -3.924 p.p. (↓23.9%), and a new assault offense by -2.604 p.p. (↓37.7%). The crime reducing effects of one year of incarceration persist even eight years after conviction. Appendix Table A.3 reports estimates for reoffending within eight years from conviction and documents a permanent reduction in ever committing a new offense -4.656 p.p. (↓7.86%) or being reincarcerated -9.8 p.p. (↓17.8%).

Estimates that include probation revocations in our measure of reoffending produce substantially more negative effects of incarceration. To illustrate this point, the first column of Table 3 reports 2SLS estimates on an indicator for being reincarcerated over the three year period after conviction (excluding any initial spell). These estimates are 50% more negative than those on committing any new offense, showing a 35.9% reduction as a result of a one-year incarceration term. The large difference between this estimate and Column 2, which repeats the final estimate from Table 2, illustrates the importance of carefully accounting for probation revocations that do not result in a new offense but nevertheless lead to reincarceration.[26] In Appendix E, we discuss several solutions for the potential bias introduced by probation revocations not associated with new criminal offenses. We present non-parametric and informative bounds that also show incarceration has crime reducing effects. The appendix also includes estimates under an independent risks

periods of time post-conviction in which the instrument stop being predictive of incarceration status was first proposed by Bhuller et al. (2018a), who study incarceration and recidivism in Norway.

[26] We examine the type of violations causing these probation revocations using 2SLS estimates of incarceration effects on different types of violations in Appendix Table A.4. The violations capture a variety of behaviors ranging from a new crime violation to a drug or a technical violation. There is no evidence that one specific violation type is driving the differences between effects on reincarceration and effects on any new offense.

assumption that show even larger crime reducing effects than those documented above.

Finally, the 2SLS estimates so far are based on a single endogenous variable, incarceration length. As we discuss in the conceptual framework section above, these estimates non-parametrically identify a weighted average of local average treatment effects. However, it is also important to examine whether the simple model of a constant, linear effect fits the data. For instance, this model implies that the effect of additional exposure to incarceration is the same from 0 to 5 months than from 15 to 20 months. The J-statistics from a Sargan-Hansen test, also known as a J-test, of treatment effect heterogeneity in Table 2 indicate that for some types of offenses (e.g., reincarceration) this simple model is rejected; however, for other offenses we cannot reject the model (e.g., any new offense). Estimates from conviction are inherently non-linear since the impacts of an additional month of incarceration on reoffending within three years of sentencing are zero for an offender who is currently serving a four-year sentence. Appendix Table A.5 reports 2SLS estimates for reoffending within one year, where it is clear that the J-statistics are significant for almost all the offense types and the null of constant and linear effects is strongly rejected by the data. This indicates the importance of treatment effect non-linearity and heterogeneity, especially when examining dynamics and effects across several horizons, which is a core motivation for the analysis in Section 7.

5.3 Treatment effect heterogeneity

To investigate the types of *offenders* driving our estimates, Appendix Table A.6 reports 2SLS estimates of incarceration length effects by the category of the defendant's initial conviction as well as the type of crime committed when they reoffend. These results show that all types of offenders are affected by incarceration. And while assault offenders are the main driver of the overall effects on new assault offenses, property and drug offenders reduce offending across all categories of crime.

Appendix Figure A.7 shows the main reduced form estimates by the following offender characteristics: previous incarceration exposure, race (black vs. non-black), and sex. There is substantial heterogeneity in the reduced form effects. Individuals without a previous incarceration spell and Caucasians experience crime reducing effects due to incarceration. For African-Americans, incarceration seems to be less rehabilitative: reductions in cumulative new offenses dissipate over a sufficiently long window (although if probation revocations are included in the recidivism measure there is still a overall reduction in crime).

Finally, treatment effect heterogeneity by felony class is discussed in Appendix H. Overall, the patterns in all classes look similar, although there is substantial variation in the shifts in incarceration exposure generated by each discontinuity. It is interesting to note that the reduced forms with the largest permanent reductions in offending also have the longest incarceration treatments. Thus, while no class shows incarceration ever increases offending post-conviction, there is some suggestive evidence that only longer sentences persistently reduce it.

5.4 Incapacitation effects and selection to incarceration

The magnitude of the incapacitation effects largely reflects the average risk of the population sentenced to incarceration. It is therefore important to examine how the risk of the compliers in our experiments compares to other populations, such as those never incarcerated. In our context, methodologies from the complier analysis literature (Imbens and Rubin, 1997; Abadie, 2002) can be used to identify the failure (reoffending) function under the no incarceration treatment for a subset of the compliers—individuals shifted to incarceration due to the instruments and whose counterfactual was therefore probation. We can identify $\mathbb{E}\left[Y_{i,t}(0)\right]$ for this group at every t using the following result

$$\frac{\mathbb{E}\left[Y_{i,t}\cdot(1-1(D_i>0))\,|Z_i=1\right]-\mathbb{E}\left[Y_{i,t}\cdot(1-1(D_i>0))\,|Z_i=0\right]}{\mathbb{E}\left[1-1(D_i>0)|Z_i=1\right]-\mathbb{E}\left[1-1(D_i>0)|Z_i=0\right]}=\mathbb{E}\left[Y_{i,t}(0)|D_i(1)>0=D_i(0)\right] \quad (6)$$

Figure 7 shows estimates of $\mathbb{E}\left[Y_{i,t}(0)|D_i(1)>0=D_i(0)\right]$ (i.e., the compliers) and $\mathbb{E}\left[Y_{i,t}|D_i=0\right]$ using different measures of reoffending such as committing any new offense, committing a new assault offense, or being reincarcerated. The results clearly show that individuals shifted to incarceration due to the instruments have higher likelihoods of criminal involvement than the average non-incarcerated individual. For example, compliers are twice as likely to commit an assault offense within one year under the probation regime. These complier rates of reoffending provide a rough estimate of crime averted due to incarceration.

In Appendix J, we extend the above analysis and discuss how, under some assumptions about treatment effect heterogeneity across complier groups, Equation (6) can be used to derive a non-parametric decomposition of IV estimates from conviction into an upper bound on the incapacitation effects and a "residual" term that can be attributed to non-incapacitation channels.[27] The decomposition highlights two key results. First, incarceration has large crime reducing impacts through incapacitation with no indication of any criminogenic effects. Second, the measure we use to assess reoffending (e.g., new offense, reincarceration) can substantially influence the magnitude of our estimates.

5.5 Estimates from "at-risk"

In this section, we turn to "at-risk" estimates in a first attempt to separate incapacitation from effects on behavior after release. This approach measures reoffending since each individual's "at-

[27]Note, that under a compliers comparability condition analogous to that in Mountjoy (2018) the incapacitation channel becomes point identified.

risk date," which is the latter of conviction and release from incarceration. For individuals who do not get an initial incarceration sentence, this measure is thus identical to the one used in the previous section. For individuals who are sentenced to incarceration, measurement starts on the day of release. Any differences in offending between the two groups using this measure are thus *not* due to differences in incapacitation resulting from the initial sentence.

Measuring reoffending from at-risk is complicated by the impact of probation revocations on the population not given an initial incarceration sentence. While many revocations occur because of new criminal activity—and thus should be properly considered reoffending—other revocations occur because of technical violations such as failing alcohol or drug test, missing a check-in with a probation officer, or traveling out of the county without authorization. When violations occur, offenders are usually incarcerated. This censors our measure of how long it takes these individuals to commit a new non-revocation offense. By construction, only individuals *not* initially given an incarceration sentence are put on probation and are thus subject to such violations. A simple comparison of times to commit a new offense between treated and untreated groups in this setting would be misleading, since means in the untreated group are measured net of the effects of probation revocation-induced incapacitation.

Table 4 reports 2SLS estimates of the effects of incarceration length (in months) on reoffending within three years of at-risk using different measures of reoffending. The estimates show that a one-year incarceration spell has an almost zero effect on committing a new offense within three years of release (-0.0383 p.p. or ↓1.08%). Although this estimate is not statistically significant, we can reject increases of more than 3.81% in the likelihood of a new offense. Since including probation revocations has a large impact on measured reoffending, we also report results for any reincarceration within three years of release in Column 1 of Table 4. One year of incarceration causes a large reduction of 8.9 p.p. (↓19%) in the likelihood of reincarceration within three years of at-risk. In Appendix L, we present reduced form results similar to those in Figure 6 that explore the dynamic effects of the discontinuities and confirm the patterns shown by the 2SLS estimates.

Since probation revocations clearly have a substantive impact on our results, we also present estimates that assume the risk of a probation revocation and committing a new criminal offense are independent. Since in reality revocation is likely to be positively correlated with criminal risk, these estimates are more conservative than those that include only new offenses. To implement the independent risks assumption, we simply drop individuals who are revoked prior to committing a new offense. The estimates in Appendix Table L.1 show reductions in the likelihood of committing any new offense by 3.68 p.p. (↓8.2%) or being reincarcerated by 3.93 p.p. (↓12.5%). However, there are no significant effects on specific types of reoffending (e.g., assaults, drug, property). In Section 7.6 (and Appendix M.5) we discuss other approaches to this problem using non-parametric bounds and sample selection corrections.

6 Threats to identification and robustness checks

In what follows, we discuss the robustness of the design with respect to two important concerns. These analyses are complementary to the validity checks we showed in Section 4.2; the combination of both re-enforces the causal interpretation of our estimates.

6.1 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 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 estimates of incarceration effects on reoffending using the offense class of each individual's most severe charge at *arraignment* and most severe *charge* brought at any point in the case, instead of the most severe *conviction*, to define the instruments. 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 >95% of cases. Thus arraigned charges are unlikely to be affected by plea negotiation. Appendix Table K.1 presents the estimates for our main results. It is clear that using the arraigned offense class yields very similar results to using the convicted offense class, 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 15% to 25% smaller. In Appendix K.1, we discuss another test that compares the characteristics of individuals who take a plea to those who do not and also shows no evidence of manipulation through plea bargaining.

6.2 Differences in the likelihood of detection

Individuals on probation may face a more intensive supervision regime, implying criminal activity will be detected more often than for those initially sentenced to incarceration after their release. 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.

First, we show that our results remain the same when using only discontinuities that do *not* cause an extensive margin shift in incarceration and generate only an intensive margin shift to longer terms of incarceration. 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. Appendix Figure K.3 presents the core reduced form estimates on reoffending using this variation, with numerical 2SLS results for different outcomes in Appendix Tables K.2 and K.3. These results show that this instrument set produces estimates with a similar sign and magnitude to our core results. The estimates also do not change when including probation revocations as reoffending. This is reassuring, since this variation implicitly compares two individuals sentenced to different lengths of incarceration as opposed to some incarceration vs. probation.

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 felony 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. Appendix Figure K.4 documents the first stage effects on the probation regime (first row) and shows there is no effect on the likelihood to reoffend, commit a new offense or be reincarcerated within three years of conviction. Appendix Figure K.5 shows this discontinuity has no effects on any pre-treatment 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 (Appendix Figure K.5, upper left corner). These findings on the effects of increased supervision intensity are in line with other studies in the literature.[28]

Overall, both of these pieces of evidence reveal that our estimated effects (in previous sections) are likely *not* driven by difference in detection probabilities and instead reflect the causal effects of incarceration on offending itself.

7 A model of the behavioral effects of incarceration

Thus far, our analysis has investigated reduced form and 2SLS effects of incarceration on reoffending both from conviction and from at-risk. This analysis does not take into account several key components of incarceration's effects on future criminal behavior. First, 2SLS estimates from conviction and from at-risk are parameterized by a single continuous endogenous variable (incarceration length) and thus do not identify any non-linearities in the effects of incarceration. These estimates imply, for example, that shifting an offender from zero to one year has the same impact as lengthening a five-year sentence by an additional year. Second, treatment effects are likely to be heterogeneous; however, 2SLS estimates of models that allow for non-linearity in the effects of incarceration (e.g., a polynomial in length of exposure) require assuming that no such

[28] Georgiou (2014) utilized a salient discontinuity in the level/intensity of supervision in Washington State and also found no effects on reoffending.

heterogeneity is present. Lastly, as is discussed in Section 2 (and Appendix C), 2SLS estimates from at-risk recover a mixture of the behavioral effects of incarceration and the effects of other time-varying confounders such as age and year of release. Attempts to control directly for these confounders can potentially lead to post-treatment adjustment bias (Rosenbaum, 1984).

To address these difficulties, we propose a single index generalized Roy (1951)-style selection model that describes how the latent propensity to commit crime varies with incarceration length and release date. The model enables us to separately identify incapacitation and behavioral effects while allowing for both non-linearity and treatment effect heterogeneity. Furthermore, the model parameters can be interpreted as causal effects and are not confounded by time-varying factors such as age at release.

7.1 Selection to incarceration

We begin by describing the selection process into incarceration. Assignment is based on a single latent index with components which are observed (e.g., prior record points, age) and unobserved to the econometrician. We use an ordered choice model for incarceration length assignment D_i that allows the choice thresholds C_d to depend on whether the offender is to the left or the right of a punishment discontinuity

$$D_i = d \quad \text{if} \quad \underbrace{C^l_{d-1}(Z^l_i)}_{\text{cut-offs}} \leq \underbrace{X_i'\gamma^l_0}_{\text{Observed component}} + \underbrace{\nu_i}_{\text{Unobserved component}} < C^l_d(\underbrace{Z^l_i}_{\text{Instrument}}) \tag{7}$$

where $\nu_i \sim N(0,1)$ and Z^l_i is an indicator for whether individual i is to the right or left of the punishment type discontinuity in felony class l, where $l \in \{E, F, G, H, I\}$ is the severity class of offender i's convicted charge. The model is estimated separately within each felony class. As is standard in ordered choice models, the thresholds are weakly increasing

$$\begin{aligned} C^l_{d-1}(Z^l_i) \leq C_d(Z^l_i) \quad \forall Z^l_i,\ l \\ C^l_{-1}(Z^l_i) = -\infty, \quad C^l_{\bar{D}}(Z^l_i) = \infty \quad \forall Z^l_i,\ l \end{aligned} \tag{8}$$

This model differs from a regular ordered probit model by allowing choice thresholds to depend on Z^l_i.[29] This implies that two offenders with similar observed and unobserved characteristics will face a different punishment regime depending on whether they are to the left or to the right of the discontinuity, capturing the variation that is introduced by the sentencing non-linearities in the grid.[30]

[29]Other studies using ordered choice models with thresholds that depend on covariates (or are themselves random variables) include Cameron and Heckman (1998); Carneiro et al. (2003); Greene and Hensher (2010).

[30]Another motivation for using this more flexible ordered choice model is that the standard ordered probit model assumes that the instruments generate additive shifts in the latent index. This implies that being above a punishment type discontinuity increases the likelihood of being assigned to *all* levels of incarceration. Formally, the

Vytlacil (2006) showed that when allowing the thresholds in an ordered choice model to be random variables the single index model is observationally equivalent to the LATE framework in Angrist and Imbens (1995) and does not impose any additional restrictions on the data generating process. Our formulation of the selection model differs from Vytlacil (2006) by allowing the thresholds to depend on the excluded-variables (i.e., the indicators for being above a punishment type discontinuity) and since we do not allow the thresholds to be stochastic. Overall, the model fits the data well both in and out of sample. Appendix M.2 describes several tests of model goodness of fit. For example, we show that the model can replicate non-parametric estimates of $\Pr(D_i(1) \geq d > D_i(0))$. Appendix M.1 describes the estimation procedure and lays out the maximum likelihood problem

7.2 Reoffending and selection

We next model the likelihood of reoffending within t months from conviction and its relationship to the selection into incarceration. To do so, we assume that conditional on the running variable (i.e., criminal history score) the instruments Z_i^l are assigned independently of the potential outcomes and unobservable factors governing selection:

$$Y_{i,t}(d), \nu_i \perp\!\!\!\perp Z_i | X_i \quad \forall d \tag{9}$$

where X_i includes grid controls such as felony class fixed effects and prior points.

Let the failure function for reoffending within t periods from conviction be a function of five factors: (i) the length of initial incarceration d, (ii) time at-risk to reoffend $t-d$, (iii) pre-conviction and time-invariant observables X_i, (iv) unobserved factors ν_i, and (v) time-varying controls $W_{i,d}$. The relationship between selection into incarceration and criminality is captured by allowing mean potential outcomes to depend on ν_i. Specifically, we assume that for each level of incarceration exposure $d \in \{0, 1, \ldots, \bar{D}\}$ the conditional expectation of $Y_{i,t}(d)$ is:[31]

$$\mathbb{E}\left[Y_{i,t}(d)|X_i, Z_i^l, W_{i,d}, \nu_i\right] = \underbrace{X_i'\xi_{t-d}}_{\text{Covar. pre-conviction}} + \underbrace{W_{i,d}'\eta_{t-d}}_{\text{Covar. at release}} + \underbrace{\beta_{t-d}\nu_i}_{\text{Selection}} + \underbrace{\overbrace{\theta^0_{d,t-d}}^{\text{Average effects}} + \overbrace{\theta^1_{d,t-d}\nu_i}^{\text{Heterogeneity}}}_{\text{Behavioral effects}} \tag{10}$$

where $t-d$ is the number of periods that the individual is at-risk to reoffend and $Y_{i,t}(d) = 0$ if $t-d \leq 0 \quad \forall i$. All coefficients vary by time at-risk $t-d$ to take into account decreasing reoffending

standard ordered probit model implies that $\Pr(D_i(1) \geq d > D_i(0)) > 0 \quad \forall d$, which is clearly rejected by the data in our case (see Appendix Figure A.4). By allowing the instruments to shift the thresholds themselves, we allow crossing a discontinuity to increase the likelihood of some durations (e.g., for two months or a year), but not others (e.g., for 10 years).

[31] Note that X_i does not include age at conviction, since if it included it then X_i and $W_{i,d}$ would have been co-linear with D_i. Conditioning on age at release, instead of age at conviction, removes from the incarceration effect the aging component by directly controlling for it.

hazards post-release.[32]

Equation (10) captures the many channels through which incarceration can impact reoffending. First, although the model describes mean reoffending t periods from *conviction*, the parameters govern behavior post-release. Incapacitation effects are instead captured by the assumption that $Y_{i,t}(d) = 0$ if $t - d \leq 0 \quad \forall i$.[33] Assignment to longer incarceration spells increases the period for which $t - d \leq 0$ and thus offending is zero. Second, the model allows incarceration assignment to affect the value of time-varying covariates such as age at release through $W_{i,d}$. Third, selection into incarceration is directly related to potential outcomes through β_{t-d}. Fourth, the average treatment effects of incarceration on reoffending post-release are represented by $\theta^0_{d,t-d}$, which allows each additional month behind bars d to have a different effect on reoffending within $t - d$ periods at risk. Finally, heterogeneity with respect to selection is governed by $\theta^1_{d,t-d}\nu_i$, which captures differences in the effects of incarceration across individuals with varying levels of latent criminality as in Garen (1984) and Card (1999) who study the choice of years of education.

By iterated expectations, Equation (10) can be written as:

$$\begin{aligned}\mathbb{E}\left[Y_{i,t}|X_i, Z_i^l, W_{i,d}, D_i = d\right] &= X_i'\xi_{t-d} + W_{i,d}'\eta_{t-d} \\ &\quad + \beta_{t-d}\overbrace{\lambda\left(X_i, Z_i^l, d\right)}^{\text{Control function}} + \underbrace{\theta^0_{d,t-d} + \theta^1_{d,t-d}\lambda\left(X_i, Z_i^l, d\right)}_{\text{Behavioral effects}}\end{aligned} \tag{11}$$

where $Y_{i,t}(d) = 0$ if $t - d \leq 0 \quad \forall i$ and $\lambda\left(X_i, Z_i^l, d\right) = \mathbb{E}\left[\nu_i|X_i, Z_i^l, D_i = d\right]$ is the generalized residual from the first stage, Equation (7). After fitting the first stage ordered choice model of time served, these generalized residuals are easily estimated. Equation (11) can then be estimated by a series of ordinary least squares regressions for each $t - d$ and using reoffending measures from *at-risk* as the outcome. This two-step "control function" estimator (Heckman and Robb, 1985; Meghir and Palme, 1999; Florens et al., 2008) is a variation of the two-step selection correction used by Heckman (1979).

To gain efficiency and make the model's estimates easier to summarize, we also estimate a simplified specification for the relationship between D_i and $\mathbb{E}\left[Y_{i,t}(d)|\cdot\right]$. This model uses a polynomial in D_i and an indicator for any incarceration sentence (instead of dummies for each month of exposure $\theta^0_{d,t-d}$ and $\theta^1_{d,t-d}$), implying that the effects of incarceration on reoffending within each time window can be captured by 6 parameters instead of the 106 parameters that are allowed in the fully general model described above. We show that this more parsimonious model still provides a good fit to the data and can replicate the experimental variation produced by the instrumental variables.

[32]Appendix Figure A.1 documents that the reoffending hazards post-release are indeed decreasing.

[33]Empirically, being charged with new crimes while in prison is exceptionally rare

7.3 Identification

Identification overall relies on the assumption that the model parameters are additively separable. This restriction is similar to assumptions in commonly used research designs such as difference-in-differences with time-varying controls.[34] Identification of $\theta^1_{d,t-d}$ specifically relies on variation in Z^l_i given $X_i = x$, $W_{i,d} = w$, $D_i = d$, and t:

$$\begin{aligned}
&\mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,d} = w, D_i = d, Z^l_i = 1\right] - \mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,d} = w, D_i = d, Z^l_i = 0\right] \qquad (12)\\
&\quad = (\theta^1_{d,t-d} + \beta_{t-d})(\lambda(x,1,d) - \lambda(x,0,d))\\
&\Rightarrow \theta^1_{d,t-d} + \beta_{t-d} = \frac{\mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,t} = w, D_i = d, Z^l_i = 1\right] - \mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,t} = w, D_i = d, Z^l_i = 0\right]}{\lambda(x,1,d) - \lambda(x,0,d)}
\end{aligned}$$

while β_{t-d} is identified by the same ratio when $d = 0$, which is the omitted treatment category. The intuition behind the identification argument above is that given similar observables ($X_i = x$ and $W_{i,d} = w$), two individuals on either side of a discontinuity who both received the same incarceration spell must differ on unobservable characteristics (i.e., ν_i). Since the offender to the right of the discontinuity faces a harsher sentencing regime, then if the two individuals received the same sentence it implies that the individual to the left is "worse" in his unobserved characteristics that are represented in our model by ν_i. This provides implicit variation in the control function term $\lambda(x, z, d)$ that is used to pin down $\theta^1_{d,t-d}$.

Another interpretation of Equation (12) is as an infeasible 2SLS regression of $Y_{i,t}$ on $\lambda(\cdot)$ using Z^l_i as an instrument. The estimator exists whenever the the denominator $\lambda(x,1,d) - \lambda(x,0,d)$ is non-zero, which is the same as requiring a sufficiently strong first-stage between $\lambda(\cdot)$ and Z^l_i. Kline and Walters (2016) use similar arguments for identification in a scenario with multiple unordered treatments and a single binary IV. Notice that since we have five binary instruments we can potentially allow the model to include interactions of X_i (or $W_{i,d}$) and ν_i.

7.4 Model estimates

We first discuss estimates of the simplified version of Equation (11) and afterwards show that the more general model with dummies for each month of incarceration exposure yields similar results. The outcome of interest is any reoffending within three years from the date the individual is back in the community and at-risk to reoffend. To circumvent issues of differential censoring due to technical probation revocations, we define reoffending as committing any new offense or probation revocation, which is practically equivalent to using reincarceration as the outcome.[35]

[34] Note that since our design uses multiple instruments, the additive separability assumption can potentially be relaxed by interacting $W_{i,d}$ with indicators for incarceration length.

[35] We also examine the the robustness of this decision by showing that estimates under an independent risks assumption are similar. In addition, the model can be extended to include an additional correction for being censored due to a technical probation revocation. Correcting this censoring problem allows us to overcome bias due to competing risks using a second control function, as discussed below.

Table 5 shows the main estimates of the simplified version of Equation (11). The coefficient on the unobserved characteristic ν_i is positive, large in magnitude and statistically significant. This indicates that individuals assigned to longer incarceration spells based on unobservable factors are also more likely to reoffend. The bottom panel shows the marginal effects of a year of incarceration on reoffending. The model estimates show that controlling for selection on unobservables using the control function shrinks substantially the marginal effect calculated using OLS. A transition from zero to one year of exposure to incarceration generates a 23% reduction in the likelihood of reoffending within three years (Column 4). However, the marginal effects are diminishing in the length of incarceration. A transition between two to three years of incarceration has an almost zero impact on reoffending.[36]

Behavioral effects for "average compliers" can be approximated by plugging estimates of the average value of ν for compliers (i.e., $\sum_{d=1}^{\bar{D}} E[\nu_i|D_i(1) \geq d > D_i(0)] \cdot \omega_d$) at each of the discontinuities into the model. Table 6 examines heterogeneity with respect to ν in the marginal effects of incarceration (in %). The treatment effects on reoffending within one year are broadly similar across compliers from different felony classes, however, effects on three year reoffending show more heterogeneity. Class I offenses (the least severe) have the largest crime reducing effects (40.8%). In addition, Table 6 documents clear patterns of non-linearity in the impacts of incarceration. Across all types of compliers (Columns 2-6) the first year of incarceration has substantial rehabilitative effects (roughly 20-28% reductions); however, lengthening an incarceration sentence from two to three years has a negligible effect on future criminality.

Next, we document the dynamics of the non-linearity in the effects of incarceration for Class I (the least severe offenses) and class E (the most severe offenses) compliers. Similar to Table 6, we report effects for "average compliers." Figure 8 shows the effects of different incarceration spells on criminal behavior post-release. There is a clear pattern of non-linearity in the impacts of incarceration. The largest rehabilitative gains are for shorter sentences and the marginal effects are quickly diminishing. Long sentences of incarceration can potentially have also marginal criminogenic effects, i.e., the marginal effects of an additional month of incarceration can be crime *increasing* for long sentences.

Finally, Appendix Table A.7 compares our preferred control function specification (Column 4 of Table 5) to IV estimates from at-risk. Columns 1 and 2 report estimates of an IV specification that models the effects of exposure to incarceration using only a single variable, the length of incarceration. The estimates show a reduction of roughly 11% in the likelihood of committing any new offense or probation revocation within three years of release for every additional year of incarceration prior to being released. This model specification does not allow for any non-linearities

[36] These decreasing effects are clearly visible also in Appendix Figure A.8, which plots the effects of various lengths of incarceration on reoffending at each $t - d$ window using estimates of the more flexible control function specification described in Equation (11). While exposure to incarceration reduces the likelihood of reoffending, the marginal effects are clearly diminishing in length of incarceration—similar to the estimates using the simplified model specification.

in the impacts of exposure to incarceration and is rejected by the data (J-test $p < 0.0001$). Column 3 allows the effects to be non-linear, but shuts down any treatment effect heterogeneity. This model fits the data substantially better and presents estimates of effect non-linearity that are similar to the control function based estimates; however, this model does not allow for any heterogeneity across either the individuals or the discontinuities we study. Column 4 reports the control function estimates, which show that treatment effect heterogeneity with respect to unobserved factors plays an important role. This result is also evident from the control function estimates in Table 6 and Figure 8.

7.4.1 Characterizing selection to incarceration

Our model can also be used to examine the selection process into incarceration. Whereas in other settings selection is informative about individuals' costs of treatment take-up, in our context the assignment process describes the considerations that motivate judges. The relationship between incarceration length and unobserved criminality follows directly from the sign of the control function ($\hat{\nu}_i$) coefficient in Table 5. The coefficient is positive and significant, indicating that incarceration for longer terms is correlated with an individual's unobserved criminality. That is, judges and prosecutors seek to incarcerate for longer durations offenders who are more likely to reoffend.

We also examine whether there is evidence of selection on gains. Appendix Figure A.9 documents a negative correlation between $\mathbb{E}\left[Y_{i,36}(36) - Y_{i,36}(0)\right]$ and D_i. This relationship arises because being incarcerated for a longer term is associated with having unobservable characteristics ν_i that predict greater behavioral gains. A similar negative correlation exists for changes only in the intensive margin between being incarcerated to 36 relative to 12 months ($\mathbb{E}\left[Y_{i,36}(36) - Y_{i,36}(12)\right]$ and D_i). Judges and prosecutors also, therefore, seek longer sentences for those most likely to desist from crime as a result of exposure to incarceration.

7.5 Replication and decomposition of reduced forms

We now use the model estimates to re-visit the RD estimates of incarceration effects on reoffending from *conviction*. First, we validate that our model can reproduce the reduced forms for each of the five punishment type discontinuities, i.e., $\mathbb{E}\left[Y_{i,t}|Z_i^l = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i^l = 0\right]$.[37]

Appendix Figure A.10 plots the non-parametric RD estimates (y-axis) for reoffending within 1, 2, 3, 4, and 5 years from conviction for each of the five felony classes (25 estimates overall) against their replications using the model estimates (x-axis) from Equation (11). If the model perfectly replicates the quasi-experimental RD estimates, then we would expect to see $R^2 = 1$ and a slope coefficient of 1. The figure shows that the simplified specification of the model matches

[37] Kline and Walters (2018) advocate using a validation exercise of this type when using a control function approach.

the RD estimates well, $R^2 = 0.972$ and the slope coefficient is 0.87 with a standard error of 0.049. Using the more general specification of the model increases the number of parameters capturing behavioral responses from 6 to 106, but only marginally increases the fit to $R^2 = 0.975$ and a slope of 0.908 (see Appendix Figure A.11).[38] Moreover, the minor deviations from the non-parametric estimates do not appear to be systematically correlated with the instrument used or the size of the reduced form effect. This leads us to conclude that the selection model approximates well the experimental variation introduced by the sentencing discontinuities.

To assess what share of the reduced form RD estimates can be explained by incapacitation relative to behavioral responses, we replicate the reduced form effects both under the null hypothesis of no behavioral responses and without imposing any restrictions, i.e., allowing for behavioral effects. The difference between the two replications can be attributed to the behavioral channel. To illustrate how this is possible, consider the following replication of a change in one month of exposure to incarceration (d vs. $d-1$) while holding fixed the time-varying covariates $W = w$ and using the characteristics of $D_i(1) \geq d > D_i(0)$ type compliers

$$\begin{aligned}\hat{\tau}(d) &\equiv \mathbb{E}\left[Y_{i,t}(d) - Y_{i,t}(d-1) | W_{i,d}, D_i(1) \geq d > D_i(0)\right] \\ &= \underbrace{\mathbb{E}\left[X_i' | D_i(1) \geq d > D_i(0)\right](\xi_{t-d} - \xi_{t-d+1}) + w'(\eta_{t-d} - \eta_{t-d+1})}_{\text{Effect of reduction in time at-risk}} \\ &\quad + \underbrace{(\theta^0_{d,t-d} - \theta^0_{d-1,t-d+1}) + (\theta^1_{d,t-d} - \theta^1_{d-1,t-d+1})\mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0)\right]}_{\text{Total behavioral effects}}\end{aligned} \tag{13}$$

The first term captures the reduction in reoffending expected from having one less month at risk for type d compliers with covariates at release $W_{i,d} = w$. The second term captures the *total* effect of d vs. $d-1$ months of incarceration on criminal behavior post release. Next we assume the null hypothesis of no behavioral effects is true and calculate $\hat{\tau}^{null}(d)$. This null implies incarceration only impacts reoffending through incapacitation and time-varying covariates, and requires that $\theta^0_{d,t-d} = \theta^0_{0,t-d}$ and $\theta^1_{d,t-d} = \theta^1_{0,t-d}$ $\forall d$. In other words, under this null, individuals who served different spells of incarceration behave in the same way once released, ceteris paribus. Formally, under this null of no behavioral effects, $\hat{\tau}(d)$ simplifies to

$$\begin{aligned}\hat{\tau}^{null}(d) &= \underbrace{\mathbb{E}\left[X_i' | D_i(1) \geq d > D_i(0)\right](\xi_{t-d} - \xi_{t-d+1}) + w'(\eta_{t-d} - \eta_{t-d+1})}_{\text{Effect of reduction in time at-risk (through covariates)}} \\ &\quad + \underbrace{(\theta^0_{0,t-d} - \theta^0_{0,t-d+1}) + (\theta^1_{0,t-d} - \theta^1_{0,t-d+1})\mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0)\right]}_{\text{Effect of reduction in time at-risk under probation regime}}\end{aligned} \tag{14}$$

Our estimate of behavioral effects is given by $\hat{\tau}(d) - \hat{\tau}^{null}(d)$. To decompose the non-parametric RD estimates shown earlier, we sum these effects across complier types using the model-implied

[38] A Wald test rejects the null that both $R^2 = 1$ and the slope is 1, however, the magnitude of the errors is quite small under both specifications of the selection model.

ACR weights for each discontinuity.[39]

$$\text{Behavioral channel} = \sum_{d=1}^{\bar{D}} \left[\hat{\tau}(d) - \hat{\tau}^{null}(d)\right] \widehat{\Pr}\left(D_i(1) \geq d > D_i(0)\right) \tag{15}$$

The results of the above decomposition are presented graphically in Figure 9 for the simplified model described in Equation (11) and in Appendix Figure A.12 for the more flexible model. The green line (square marker) represents the model replication of the reduced form effects (Equation 13), the black line (diamond marker) shows the model replication under the null of *no* behavioral effects (Equation 14), and the blue line (round marker) reports the estimates of the behavioral channel (Equation 15). A similar pattern emerges across all felony classes. In the initial months after conviction, the incapacitation channel alone can explain all the reductions in reoffending. However, as time goes by, the share of the behavioral channel increases, although the crime reducing effects are also diminishing. Overall, after five years from conviction, the model shows reductions in reoffending across all the different felony classes explained primarily by the behavioral channel.[40]

These findings are summarized in Table 7. Within one year of conviction, the majority of the reduced form effects can be explained by incapacitation alone, with the behavioral channel explaining between 0.05% to 10% of the total reduction. However, within five years from conviction, the importance of the two channels is reversed, with behavioral effects now explaining between 30% to 84% of the reductions in reoffending. This exercise allows us to go re-interpret our main results in Figure 6. The control function estimates show that eight years after conviction the behavioral channel explains the majority of the observed crime-reducing effects.

Importantly, this estimate of behavioral effects in does not require extrapolating away from the individuals affected directly by the discontinuities, i.e., the compliers. While the estimates in Table 5 report average treatment effects across all individuals, these decompositions show the behavioral effects for the populations of compliers that are directly influenced by the discontinuities in punishment at each felony class.

7.6 Probation revocations as non-random censoring

As discussed earlier, probation revocations can be viewed as a competing risk for being charged or convicted of a new offense. So far, we used both new offenses and probation revocations in our outcome measure, which may overestimate the quantity of socially costly crime in the probation population. We now examine the implication of this decision by instead assuming that revocation and new offense risks are independent, allowing us to simply drop observations that have a probation revocation prior to committing a new offense and are therefore censored by the

[39]These weights are estimated using Equation (M.9) from Appendix M.2.

[40]Appendix M.4 shows also how the model estimates can be used to conduct a decomposition of a marginal increase in incarceration exposure of one month into incapacitation and behavioral channels.

competing risk.

If probation revocations are independent from new offense risk, then this sample adjustment solves the censoring problem. However, it is also possible that revocation risk and new offense risk are *positively* correlated. To allow for a more flexible dependence between revocation and reoffending risks, we also introduce a second selection model and construct a second selection correction for whether the individual had a probation revocation prior to committing a new offense. This is done by directly modeling the selection process into probation revocations and showing that our instruments provide enough variation to identify *two* control function terms. Appendix M.5 describes the extended selection model and discusses identification and estimation using a three-step control function estimator.

Appendix Table A.8 shows estimates of the simplified model assuming independent risks. Similar to before, the coefficient on the control function ($\hat{\nu}_i$) is positive and significant, indicating selection to incarceration based on the likelihood of reoffending. The OLS estimate of incarceration effects is now positive ($\uparrow$ 2.477%) and opposite in sign to both 2SLS ($\downarrow$ 13.75%) and control function ($\downarrow$ 7.09%) estimates. As before, there are diminishing marginal effects of incarceration, with the majority of the reduction coming in the first two years. Appendix Table A.9 examines heterogeneity of the marginal effects of incarceration (in %) with respect to the compliers in each class. The treatment effect estimates are similar across compliers with different average unobserved characteristics, although there is substantial differences in the effects across the different groups of compliers.

Column 6 includes the sample selection correction for censoring due to probation revocation. The coefficient on the correction term is close to zero and is not statistically significant, indicating that the selection to technical probation revocations does not seem to be correlated with likelihoods of reoffending given the other controls in the model. This results suggests that the independent risks assumption is a good approximation and does not introduce sample selection to a degree that would influence the estimated effects of incarceration exposure.

8 Policy implications

In this section we investigate some of the policy implications of estimates from both the earlier non-parametric IV analysis and the model-based estimates. We begin by describing policy implications that can be derived using only the IV analysis and then proceed to consider policy counterfactuals that use the model estimates and extrapolate beyond the local average treatment effects identified in the first part of the study.

8.1 Costs and benefits of incarceration

To summarize and quantify our estimates from the RD analysis, we conduct a simplified cost-benefit comparison of the cumulative value of crime averted by an initial incarceration spell relative to the costs of incarceration for the marginal offender. The primary difficulty in doing so is assigning dollar values to criminal events (e.g., assault, murder, DWI). We use two different and complementary approaches. The first is a "break-even" approach that asks how costly the marginal offense needs to be to justify the costs of incarceration—that is, how much society needs to value the marginal averted offense to justify the costs associated with an incarceration spell. Our break-even estimates are based on 2SLS estimates of Equation (5), where the treatment is cumulative months incarcerated up to month t (e.g., within three years), from both initial and subsequent sentences, and the outcome is cumulative reoffending. To obtain the break-even value we divide the 2SLS estimate by the cost of a month of incarceration. This break-even estimator can be thought of as

$$\text{Break-even value} \equiv \underbrace{\frac{\Delta\,(\text{Cumulative number of new offenses})}{\Delta\,(\text{Cumulative months incarcerated})}}_{\beta_{2SLS}} \times \frac{1}{\text{Incar. costs per month}} \quad (16)$$

Table 8 reports break-even estimates overall and for each offense class separately. It also includes break-even values for several measures of reoffending (e.g., any new offense, new felony offense) and for different time horizons from conviction (e.g., 1, 3, and 8 years). Lengthening the incarceration spell of an offender by one month reduces cumulative new offenses by -0.0298 after eight years and it costs roughly \$2,738 per month. Thus, the per offense break-even value is \$91,784. The associated per offense break-even value is \$41,904 when including probation revocations in our measure of cumulative reoffending and it is \$164,081 if restricting attention only to felony offenses. The break-even estimate are increasing over time from \$58,809 within one year, to \$77,142 after five years, and \$91,784 after eight years from conviction. This pattern emerges because in the initial periods the incapacitation effect dominates; however, over time the initially incarcerated offenders are released and are able to reoffend.

To visually illustrate how the break-even estimates dynamically evolve we plot the reduced form effects on both cumulative costs of incarceration (i.e., cumulative incarceration multiplied by its costs) and the effects one cumulative number of new offenses. To summarize all five discontinuities in one coefficient, we estimate Equation 4 while imposing that being above a punishment discontinuity has the same effect across felony classes.Appendix Figure A.13, Panel (a), plots effects on cumulative number of new offense (red line and right y-axis) and cumulative costs of incarceration (black line and left y-axis). Dividing the effect on cumulative incarceration costs (black line) by the effect on cumulative new offenses (red line) will yield the break-even value. The figure shows that after an eight year period from conviction the break-even estimates seem to be stabilized at roughly $92,000 per offense.

Importantly, note that an initial incarceration sentence influences the cost-benefit estimates in two ways. First, it has a crime reducing effect. Second, it reduces the likelihood of a future incarceration spell. Thus, estimates that use only the length of the initial incarceration spell as the treatment of interest in the cost-benefit analysis will underestimate the costs associated with probation, since many offenders initially sentenced to probation get reincarcerated.

The second approach is to use estimates of the costs of different types of criminal events. Since there is large variation in the literature on how to value different types of crimes, we use upper and lower bounds for each crime category (see Appendix Table A.10 for common estimates in the literature). Appendix Figure A.13, Panel (b), shows that in the short run—when the incapacitation effect dominate—the value of crime averted can outweigh the costs. However, as incarcerated individuals are released, the difference in the value of crime averted falls while the differences in the costs of incarceration remain stable. Eight years after conviction, the costs of incarceration can outweigh the gains.

A few important caveats are in order. First, our estimates do not take into account the disutility of incarceration for offenders themselves (nor any potential direct utility for victims). Indeed, we only compare the value of crime averted to the costs of incarceration. Second, deterrence effects are not taken into account. Lastly, other social costs such as the opportunity costs of lost earnings or spillovers onto defendants' families and communities are also ignored.

8.2 Extrapolating beyond the discontinuities

In this section we examine the optimality of the current sentencing guidelines and present suggestive evidence that there is potential for Pareto improvements. Our results so far have shown that assignment to incarceration is correlated with both selection on levels (reoffending probabilities if not incarcerated) and selection on gains from exposure to incarceration. These patterns are encouraging, but they do not imply that the current system is optimal. Figure 10 plots the share of offenders who are incarcerated (blue line, right y-axis) and the density of reoffending probabilities if not incarcerated predicted using the model estimates (black line, left y-axis). The figure confirms that there is selection on levels. However, many offenders with low likelihoods of reoffending are currently incarcerated, which suggests that other sentencing guidelines/regimes might be able to achieve an increase in public safety combined with a lower incarceration rate. It is beyond the scope of this paper to derive the optimal level of incarceration based on observable (and expected unobservable) characteristics, but the above presents suggestive evidence that the current system is not optimal.

9 Concluding remarks

Our analysis shows that incarceration substantially reduces crime in the years after conviction relative to a counterfactual of probation, i.e., community supervision. The effects are not concentrated among a specific type of criminal incident: we observe reductions in violent crime, property crime, and reincarceration events. We then estimate a semi-parametric model for the treatment effects of incarceration that shows that although the majority of short-run effects are explained by incapacitation, incarceration also moderately reduces offending after release. To summarize our estimates, we conduct a simplified cost-benefit analysis which suggests that, despite the reductions in crime from more aggressive sentencing, the high costs of incarceration may outweigh the social benefits of lower crime.

Our estimates are an important contribution to the on going debate over U.S. 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.[41] While our estimates show that incarceration sentences do not make offenders more likely to offend, they also demonstrate that incarceration has room to rehabilitate inmates further, especially when compared to carceral regimes in other developed countries. 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.

Similarly, since any crime-reducing effect of incarceration is measured relative to a probation counterfactual, this implies that investments and reform in the probation system are necessary to reduce incarceration rates without increasing crime. Lastly, we show that, on the margin, increased monitoring does not reduce reoffending among probationers. This suggests reform efforts need to be directed towards measures that can rehabilitate offenders and decrease the relative attractiveness of crime—such as job training programs—among the probation population.

[41] See, for example, Attorney General Jeff Sessions 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

Abadie, Alberto, "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 2002, pp. 284–293.

Agan, Amanda and Sonja Starr, "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *The Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

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

Aneja, Abhay and Carlos Fernando Avenancio-León, "Credit-Driven Crime Cycles: The Connection Between Incarceration and Access to Credit," 2018.

Angrist, Joshua D. and Guido W. Imbens, "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 1995, *90* (430), 431–442.

_ , _ , and Donald B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, *91* (434), 444–455.

Arnold, David, Will Dobbie, and Crystal S Yang, "Racial bias in bail decisions," *The Quarterly Journal of Economics*, 2018, *133* (4), 1885–1932.

Arteaga, Carolina, "The Cost of Bad Parents: Evidence from the Effects of Incarceration on Children's Education," 2018. Working paper.

Austin, James and Robert Lawson, "Assessment of California Parole Violations and Recommended Intermediate Programs and Policies," Technical Report, San Francisco: National Council on Crime and Delinquency 1998.

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

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

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad, "Incarceration, Recidivism and Employment," Working Paper 22648, National Bureau of Economic Research 2018.

_ , _ , _ , and _ , "Intergenerational Effects of Incarceration," *AEA Papers and Proceedings*, 2018, *108*, 234–40.

Bowers, Jake. and Ben B. Hansen, "Attributing Effects to A Cluster Randomized Get-Out-The-Vote Campaign," *Journal of the American Statistical Association*, 2009, *104* (487), 873–885.

Buonanno, Paolo and Steven Raphael, "Incarceration and Incapacitation: Evidence from the 2006 Italian Collective Pardon," *The American Economic Review*, 2013, *103* (6), 2437–2465.

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

Card, David, "The causal effect of education on earnings," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, 1 ed., Vol. 3, Part A, Elsevier, 1999, chapter 30, pp. 1801–1863.

_ , David S. Lee, Zhuan Pei, and Andrea Weber, "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 2015, *83* (6), 2453–2483.

Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman, "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," *International Economic Review*, 2003, *44* (2), 361–422.

Chalfin, Aaron and Justin McCrary, "Are U.S. Cities Underpoliced? Theory and Evidence," *Review of Economics and Statistics*, 2017.

Chen, M Keith and Jesse M Shapiro, "Do harsher prison conditions reduce recidivism? A discontinuity-based approach," *American Law and Economics Review*, 2007, *9* (1), 1–29.

Clark, Damon and Emilia Del Bono, "The long-run effects of attending an elite school: Evidence from the united kingdom," *American Economic Journal: Applied Economics*, 2016, *8* (1), 150–76.

Cohen, Mark A., Roland T. Rust, Sara Steen, and Simon T. Tidd, "WILLINGNESS-TO-PAY FOR CRIME CONTROL PROGRAMS," *Criminology*, 2004, *42* (1), 89–110.

Dobbie, Will, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks, "The intergenerational effects of parental incarceration," Technical Report, National Bureau of Economic Research 2018.

_ , Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," Technical Report 2 February 2018.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova, "The Deterrent Effects of Prison: Evidence from a Natural Experiment," *Journal of Political Economy*, 2009, *117* (2), 257–280.

Eberwein, Curtis, John C. Ham, and Robert J. Lalonde, "The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data," *The Review of Economic Studies*, 1997, *64* (4), 655–682.

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

Florens, Jean-Pierre, James J Heckman, Costas Meghir, and Edward Vytlacil, "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects," *Econometrica*, 2008, *76* (5), 1191–1206.

Franco, Catalina, David J. Harding, Jeffrey Morenoff, and Shawn D. Bushway, "Estimating the Effect of Imprisonment on Recidivism: Evidence from a Regression Discontinuity Design," 2017. Accessed 9/30/2018 from https://catalinafranco.com/research/.

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

Garen, John, "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 1984, *52* (5), 1199–1218.

Georgiou, Georgios, "Does increased post-release supervision of criminal offenders reduce recidivism? Evidence from a statewide quasi-experiment," *International Review of Law and Economics*, 2014, *37*, 221–243.

Gourieroux, Christian, Alain Monfort, Eric Renault, and Alain Trognon, "Generalised residuals," *Journal of Econometrics*, 1987, *34* (1), 5 – 32.

Green, Donald P. and Daniel Winik, "Using Random Judge Assignments to Estimate the Effects of Incarceration and Probation on Recidivism Among Drug Offenders," *Criminology*, 2010, *48* (2), 357–387.

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

Grogger, Jeffrey, "The Effect of Arrests on the Employment and Earnings of Young Men," *The Quarterly Journal of Economics*, 1995, *110* (1), 51–71.

Ham, John and Robert LaLonde, "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 1996, *64* (1), 175–205.

Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway, "Short- and long-term effects of imprisonment on future felony convictions and prison admissions," *Proceedings of the National Academy of Sciences*, 2017, *114* (42), 11103–11108.

_ , _ , _ , and _ , "Imprisonment and Labor Market Outcomes: Evidence from a Natural Experiment," *American Journal of Sociology*, 2018, *124* (1), 49–110.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

_ and Edward E. Leamer, eds, "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 James J. Heckman and Edward E. Leamer, eds., , Elsevier, 2007.

_ and Edward J. Vytlacil, "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in James J. Heckman and Edward E. Leamer, eds., , Vol. 6, Part B of *Handbook of Econometrics*, Elsevier, 2007, pp. 4779 – 4874.

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

Heckman, James J and Richard Robb, "Alternative methods for evaluating the impact of interventions: An overview," *Journal of econometrics*, 1985, *30* (1-2), 239–267.

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

Huttunen, K., M. Kaila, M. Kaila, T. Kosonen, and E. Nix, "Shared Punishment? The Impact of Incarcerating Fathers on Child Outcomes," 2019.

Imbens, Guido and Donald Rubin, "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 1997, *64* (4), 555–574.

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

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

_ **and Christopher Walters**, "On Heckits, LATE, and Numerical Equivalence," 2018.

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

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

Kyckelhahn, T., "Justice Expenditures and Employment, FY 1982-2007 Statistical Tables," Report NCJ 236218, U.S. Department of Justice 2011.

Lee, David S, "Randomized experiments from non-random selection in US House elections," *Journal of Econometrics*, 2008, *142* (2), 675–697.

Levitt, Steven D., "The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation," *The Quarterly Journal of Economics*, 1996, *111* (2), 319–351.

Lochner, Lance and Enrico Moretti, "Estimating and Testing Models with Many Treatment Levels and Limited Instruments," *The Review of Economics and Statistics*, 2015, *97* (2), 387–397.

Loeffer, Charls E., "Does Imprisonment Alter Life Course? Evidence on Crime and Employment from a Natural Experiment," *Criminology*, 2 2013, *51* (1), 137–166.

Lofstrom, Magnus and Steven Raphael, "Crime, the Criminal Justice System, and Socioeconomic Inequality," *The Journal of Economic Perspectives*, 2016, *30* (2), 103–126.

Londono-Velez, Juliana, Catherine Rodriguez, and Fabio Sánchez, "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia," 2018.

Løken, Katrine V., Magne Mogstad, and Matthew Wiswall, "What Linear Estimators Miss: The Effects of Family Income on Child Outcomes," *American Economic Journal: Applied Economics*, April 2012, *4* (2), 1–35.

Marvell, Thomas B. and Carlisle E. Moody, "Prison population growth and crime reduction," *Journal of Quantitative Criminology*, Jun 1994, *10* (2), 109–140.

Maurin, Eric and Aurelie Ouss, "Sentence Reductions and Recidivism: Lessons from the Bastille Day Quasi Experiment," February 2009. IZA DP No. 3990.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

_ **and Sarath Sanga**, "General Equilibrium Effects of Prison on Crime: Evidence from International Comparisons," *Cato Papers on Public Policy*, 2012, *2.*

Mears, Daniel P., Joshua C. Cochran, William D. Bales, and Avinash S. Bhati, "Recidivism and Time Served in Prison," *Journal of Criminal Law and Criminology*, 2016, *106* (1).

Meghir, Costas and Marten Palme, "Assessing the effect of schooling on earnings using a social experiment," 1999.

Miles, Thomas J. and Jens Ludwig, "The Silence of the Lambdas: Deterring Incapacitation Research," *Journal of Quantitative Criminology*, 2007, *23* (4), 287–301.

Miller, T., M. Cohen, and B. Wiersema, "Victim costs and consequences: A new look," National Institute of Justice Research Report NCJ-155282, U.S. Department of Justice 1996.

Mountjoy, Jack, "Community colleges and upward mobility," *Unpublished manuscript*, 2018.

Mueller-Smith, Michael, "The Criminal and Labor market Impacts of Incarceration," Working Paper 2015.

Nagin, Daniel S. and G. Matthew Snodgrass, "The Effect of Incarceration on Re-Offending: Evidence from a Natural Experiment in Pennsylvania," *Journal of Quantitative Criminology*, 2013, *29* (4), 601–642.

National Center for State Courts, "State Sentencing Guidelines Profiles and Continuum," Technical Report 2008.

Norris, Samuel, "Judicial Errors: Evidence from Refugee Appeals," 2018.

_ , Matthew Pecenco, and Jeffrey Weave, "The Intergenerational and Sibling Effects of Incarceration: Evidence from Ohio," 2018.

Owens, Emily G., "More Time, Less Crime? Estimating the Incapacitative Effects of Sentence Enhancements," *Journal of Law and Economics*, 2009, pp. 551–579.

Peterson, Arthur V, "Bounds for a Joint Distribution Function With Fixed Sub-distribution Functions: Application to Competing Risks," *Proceedings of the National Academy of Sciences*, 1976, *73* (1), 11–13.

Raphael, Steven, "The New Scarlet Letter? Negotiating the U.S. Labor Market with a Criminal Record," 2014.

_ and Magnus Lofstrom, "Incarceration and Crime: Evidence from California's Realignment Sentencing Reform," 2015.

Roodman, David, "Impact of Incarceration on Crime," Technical Report 2017.

Rosenbaum, Paul R., "The Consquences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society. Series A (General)*, 1984, *147* (5), 656–666.

Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, *3* (2), 135–146.

Sanderson, Eleanor and Frank Windmeijer, "A weak instrument F-test in linear IV models with multiple endogenous variables," *Journal of Econometrics*, 2016, *190* (2), 212 – 221. Endogeneity Problems in Econometrics.

Solis, Alex, "Credit access and college enrollment," *Journal of Political Economy*, 2017, *125* (2), 562–622.

Stevenson, Megan, "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes," Working Paper 2016.

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

U.S. Department of Justice, "National Assessment of Structured Sentencing," Report, Bureau of Justice Assistance 1996.

Vytlacil, Edward, "Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *The Review of Economics and Statistics*, 2006, *88* (3), 578–581.

Wooldridge, Jeffrey M., "Control Function Methods in Applied Econometrics," *Journal of Human Resources*, 2015, *50* (2), 420–445.

Yang, Crystal S., "Local labor markets and criminal recidivism," *Journal of Public Economics*, 2017, *147*, 16 – 29.

Zapryanova, Mariyana, "The Efects of Time in Prison and Time on Parole on Recidivism," 2017.

Figures

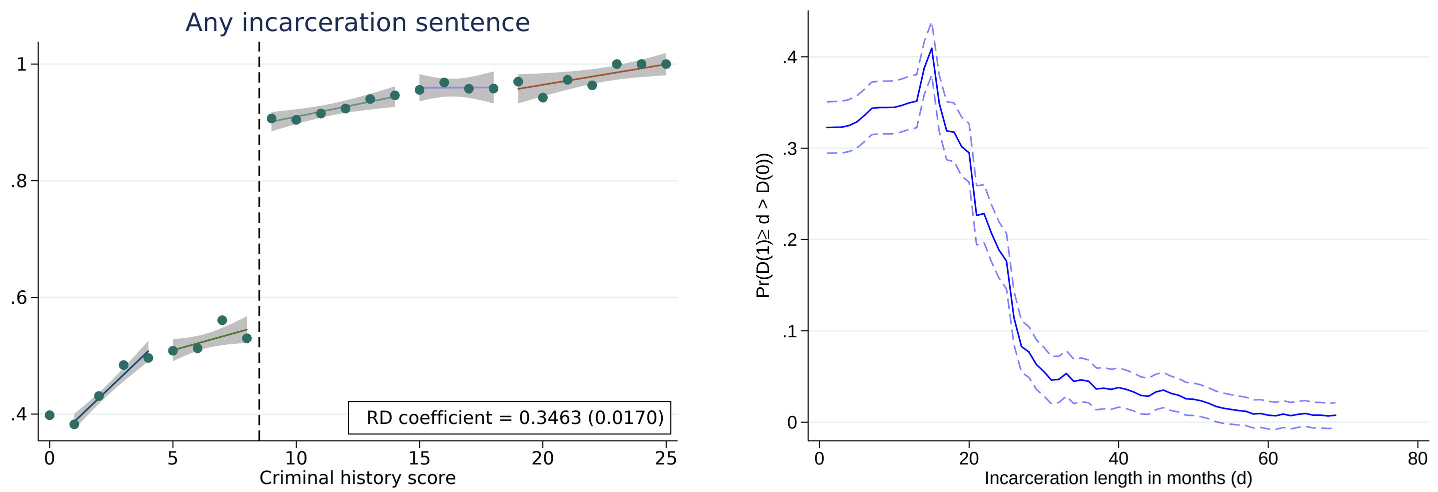
Figure 1: Example of Sentencing Guidelines

	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

Notes: This figure shows the sentencing guidelines, or "grid," applicable to offenses committed after 12/1/1995 but before 12/1/2009. Appendix D includes the full set of guidelines from 1995 to the present. The rows of the table refer to the convicted offense severity class. The columns indicate the criminal history, or "prior points", scores. Each individual is assigned a prior points score that is a weighted sum of past convictions based on severity and timing. Prior points scores are classified into prior record levels (columns) according to legislated thresholds. The numbers listed indicate minimum sentences for each offense class and prior record level combination, which we refer to as a grid "cell." The minimum sentences are specified for three different ranges: Aggravated, presumptive, and mitigated. The maximum sentence is 120% of the minimum sentence. The majority of crimes are sentenced in the presumptive range. Each grid cell is assigned a set of allowable sentence types: "A" denotes active incarceration and "C/I" denote probation, where probation of type I has more monitoring than probation of type "C" and can also involve a short prison spell. The red lines indicate places in the grid the recommended sentence type changes. Indicators for having an offense class and prior record points combinations that fall to the right of each red line are our core instruments.

Figure 2: Illustration of First Stage: Sentencing outcomes by prior points for Class F offender

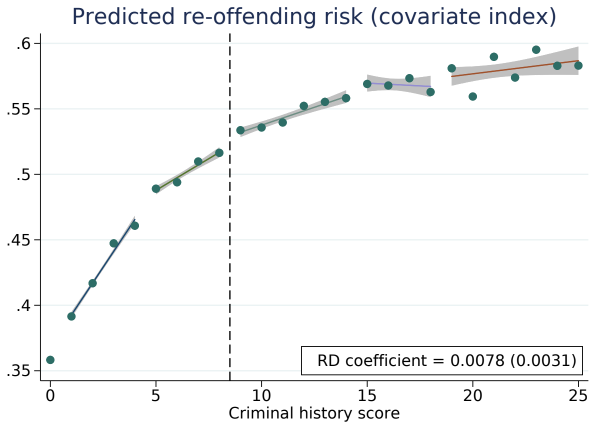


(a) Any incarceration

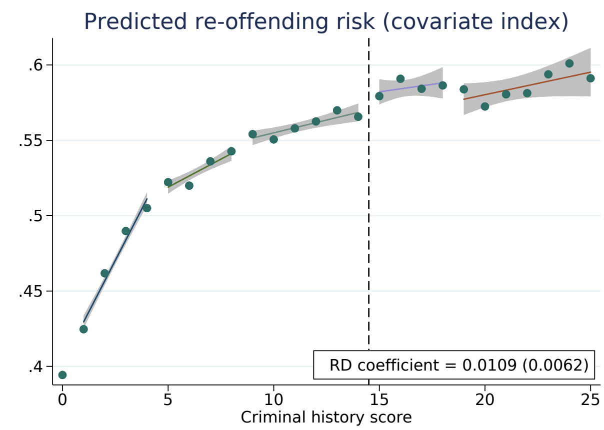
(b) Distributional effects ($\Pr(D_i(1) \geq d > D_i(0))$)

Notes: 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 Panel (a), the share of offenders sentenced to incarceration is plotted agains the running variable, prior record points. Panel (b) plots estimates of the shifts in incarceration exposure generated by the instrument, which correspond to the weights in the average causal response. These shifts reflect the probability an offender would spend less than d months incarcerated if assigned $Z_i = 0$, but at least d months if assigned $Z_i = 1$. This probability can be estimated non-parametrically as $\mathbb{E}[1(D_i \geq d)|Z_i = 1] - \mathbb{E}[1(D_i \geq d)|Z_i = 0]$, which corresponds to the coefficient on Z_i in our first stage specification when $1(D_i \geq d)$ is the outcome. Panel (a) shows only offenses sentenced under the sentencing grid that applied to offenses committed between 1995 to 2009. In 2009, the guidelines changed and the discontinuities shifted by one prior points either to the left or to the right. All official grids are in Appendix D. Similar figures for all other classes are in Appendix Figure A.3. Standard errors are clustered at the individual level.

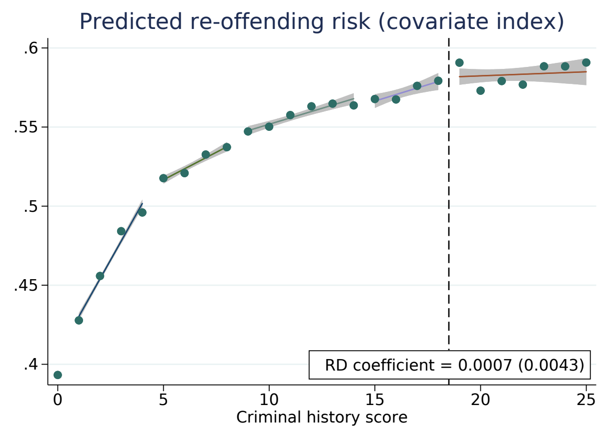
Figure 3: Predicted reoffending score by offense severity class and prior points



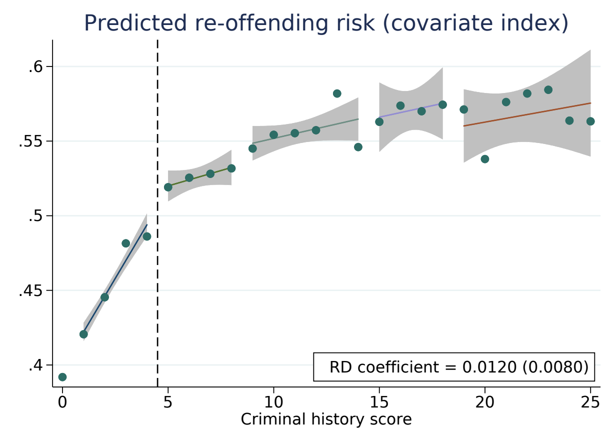
(a) Class I



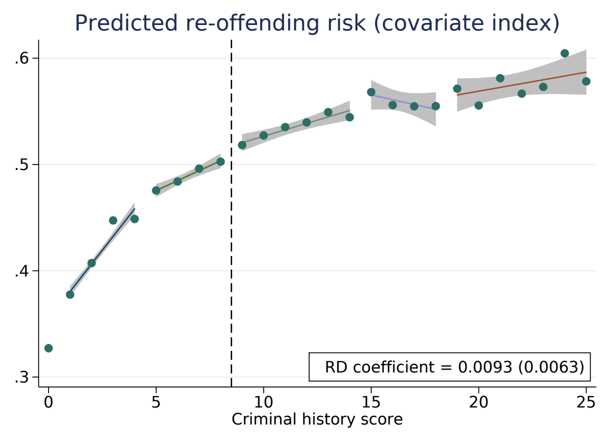
(b) Class G



(c) Class H



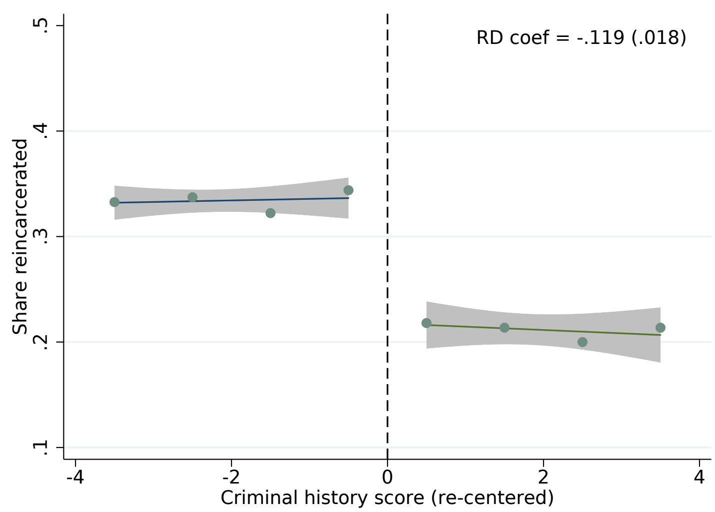
(d) Class E



(e) Class F

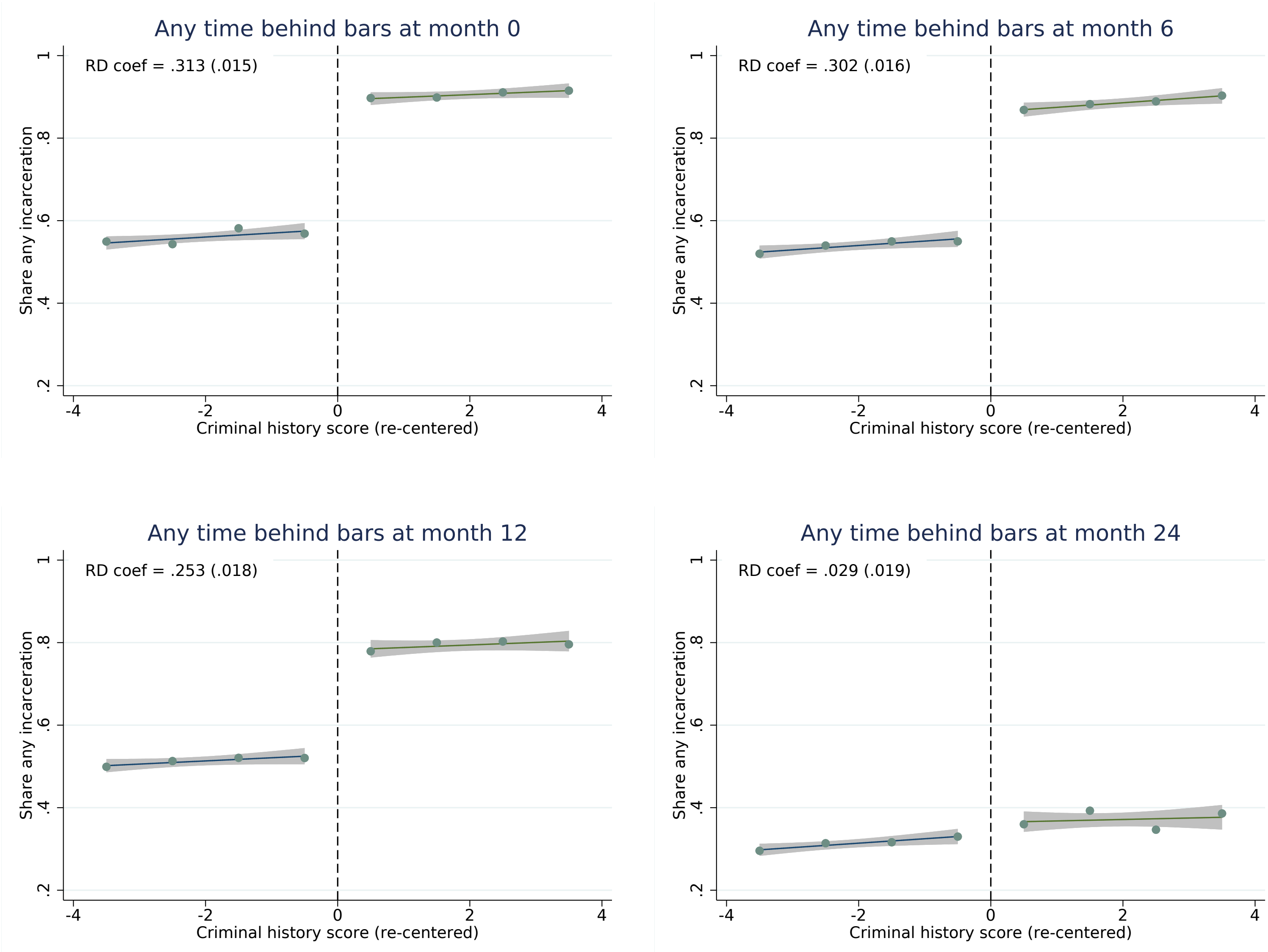
Notes: This figure demonstrates that a summary index of the covariates varies smoothly across the punishment type discontinuities in each offense class. The x-axis in all plots reports the number of prior record points. The y-axis shows mean predicted reoffending from a linear regression of all available covariates (e.g., age, race, criminal history) on reoffending within 3 years of the time of release (using only non-incarcerated offenders). We use this index because there are many potentially important pre-treatment covariates. Summarizing imbalance by the covariates' relationship to the outcome surface is a common methodology in the literature (Bowers and Hansen, 2009; Card et al., 2015; Londono-Velez et al., 2018). Standard errors are clustered at the individual level. Only offenses sentenced under the sentencing grid that applied to offenses committed between 1995 to 2009 are plotted.

Figure 4: Share reincarcerated within 3 years of conviction (Class F example)



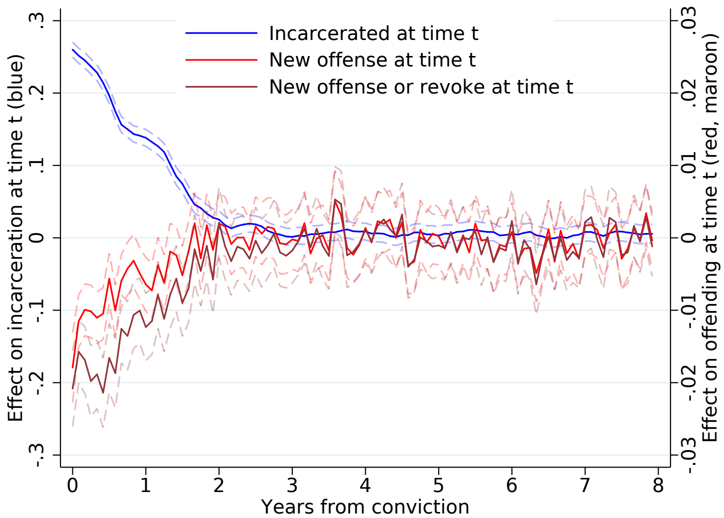
Notes: This figure shows the reduced form effect of being to the right of the punishment type discontinuity in Class F on the likelihood of reincarceration within 3 years of conviction. The x-axis shows the re-centered value of prior record points. The y-axis reports the share of individual reincarcerated within 3 years of conviction. Our parameter of interest, which is reported in the figure, is the coefficient on an indicator for whether the individual is to the right of the punishment type discontinuity. Standard errors are clustered at the individual level.

Figure 5: Dynamics of incarceration status *within* a given month from conviction (Class F example)

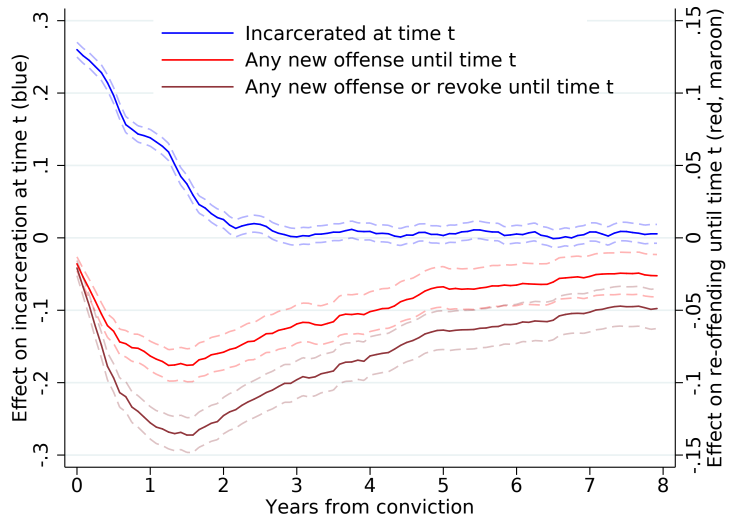


Notes: This figure shows the reduced form effect of being to the right of the punishment type discontinuity in Class F on the likelihood of being incarcerated at any point within month t after conviction. The x-axis shows the re-centered value of prior record points. The y-axis reports the share of individual who spent any time behind bars within month t. For example, the y-axis in the upper-left plot shows the share incarcerated at point in month 0. Similarly, the y-axis in the lower-right plot shows the share of offenders incarcerated for at any point in month 24 after the date of conviction. Our parameter of interest, which is reported in the figure, is the coefficient on an indicator for whether the individual is to the right of the punishment type discontinuity. Standard errors are clustered at the individual level.

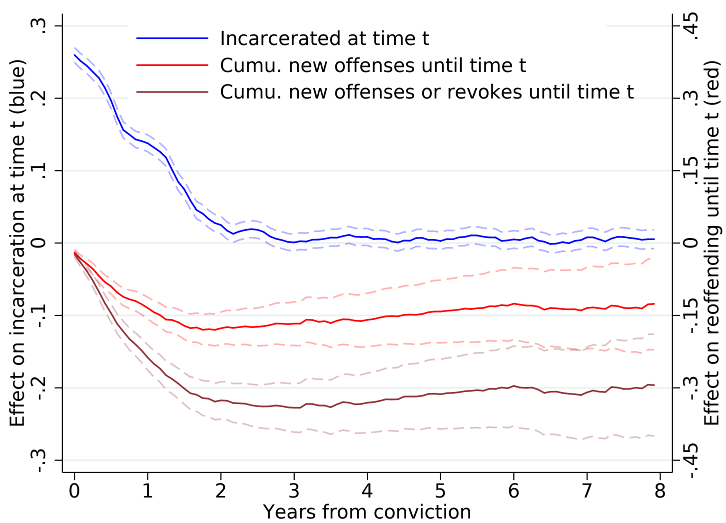
Figure 6: Effects on reoffending from conviction



(a) Reoffending in period t from conviction



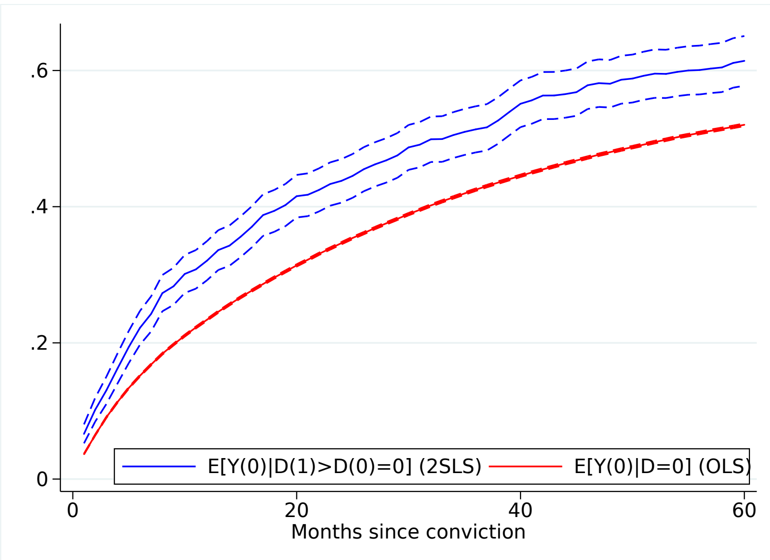
(b) Any reoffending until period t from conviction



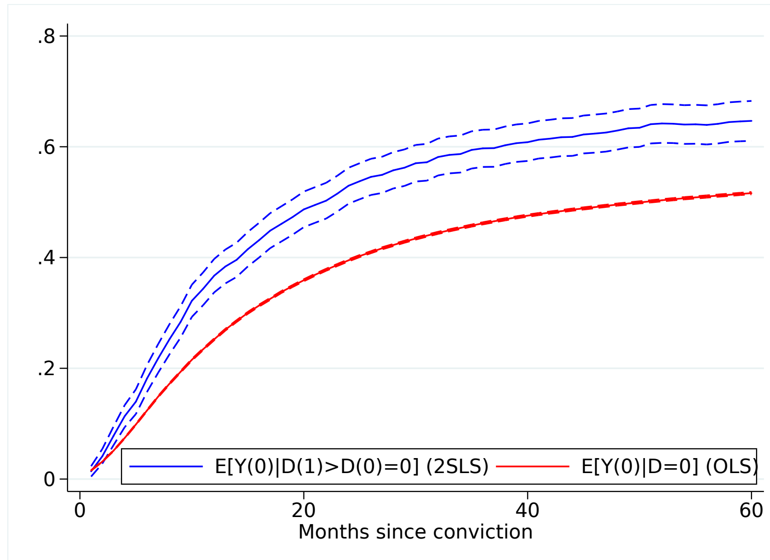
(c) Cumulative reoffending until period t

Notes: This figure shows the reduced form effects of being to the right of a punishment type discontinuity on several key outcomes. The blue line (left y-axis) in all panels shows effects on an indicator for being incarcerated at any point *in* month t from conviction. For example, the estimates in Figure 5 correspond to four points in the blue line at $t = 0$, 6, 12, and 24. In Panel (a), the red line (right y-axis) reports effects on an indicator for committing a new offense *in* month t; the maroon line (right y-axis) includes probation revocations in the outcome. In Panel (b), the red line (right y-axis) reports effects on ever committing *any* new offense *until* month t from conviction; the maroon line (right y-axis) again includes probation revocations. In Panel (c), the red line (right y-axis) reports effects on the cumulative number of new offenses committed until month t, with the maroon line including probation revocations. Standard errors are clustered by individual. Each point in each figure is an estimate of γ^{RF} for the relevant outcome for month t. This estimate is a constrained version of Equation 4 that requires the coefficients on all instruments to be the same (i.e., $\gamma^2_{E,4} = \gamma^2_{F,9} = \gamma^2_{G,14} = \gamma^2_{H,19} = \gamma^2_{I,9} = \gamma^{RF}$). This strategy averages across all five offense classes and instruments, but collapses our variation into a single coefficient. γ^{RF} can therefore be thought of as the average reduced form effect across the five punishment type discontinuities (taking the actual average of the individual reduced forms yields highly similar results). The notation used is based on the guidelines in place prior to the 2009 reform, although all observations are used in estimation. The regression specifications include as controls demographics (e.g., race, gender, age FEs), FEs for the duration of time previously incarcerated, the number of past incarceration spells and the number of past convictions, county FEs, and year FEs. Estimates without controls yield similar results (see Table 2).

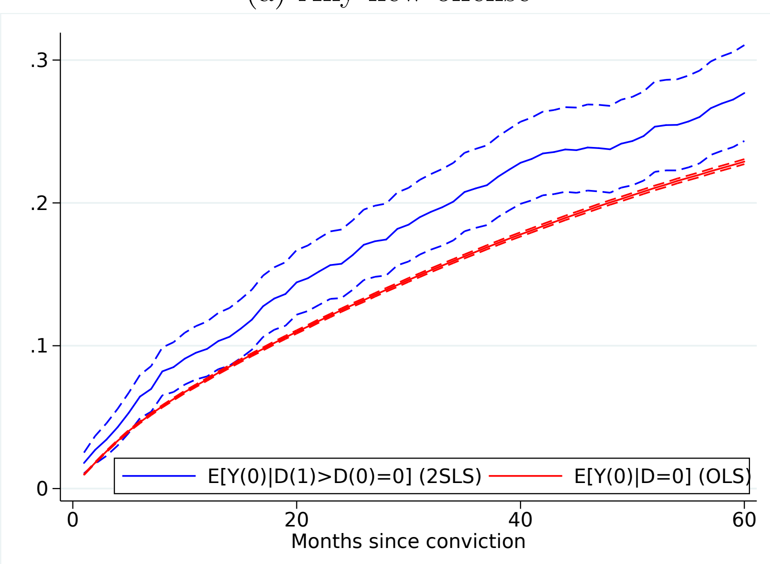
Figure 7: Reoffending patterns if not incarcerated: Compliers vs. never incarcerated



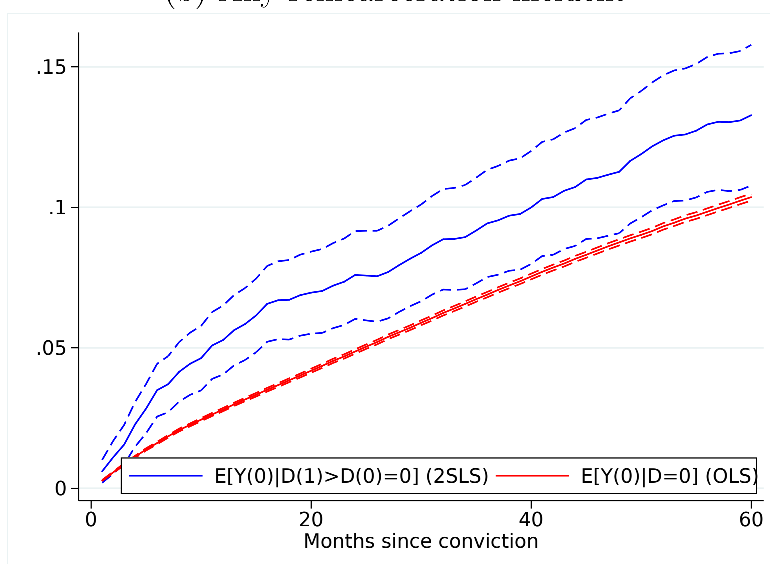
(a) Any new offense



(b) Any reincarceration incident



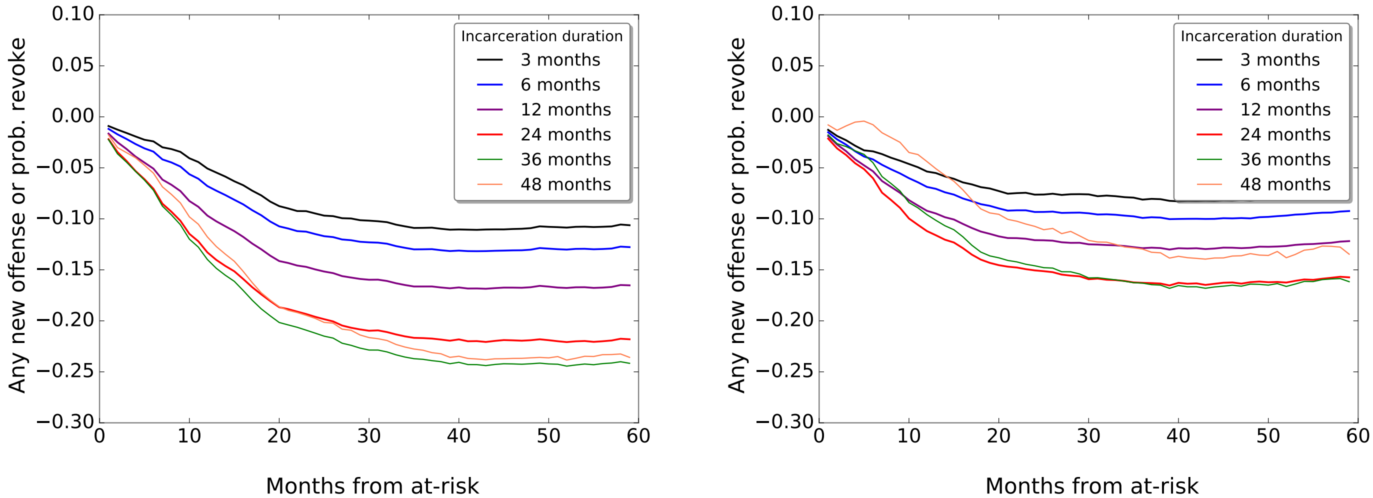
(c) Any new drug offense



(d) Any new assault offense

Notes: This figure shows estimates of the likelihood of reoffending when not incarcerated. The red line plots the reoffending probability of offenders who are not been incarcerated ($\mathbb{E}\left[Y_{i,t}(0)|D_i = 0\right]$). The blue line reports estimates for a subset of compliers who receive no incarceration when $Z_i = 0$, but some positive amount otherwise ($\mathbb{E}\left[Y_{i,t}(0)|D_i(1) > D_i(0) = 0\right]$). We can recover $\mathbb{E}\left[Y_{i,t}(0)|D_i(1) > D_i(0) = 0\right]$ using formulas derived by Imbens and Rubin (1997) and Abadie (2002), while the red line is directly observable.

Figure 8: Heterogeneity and non-linearity in the behavioral effects of incarceration

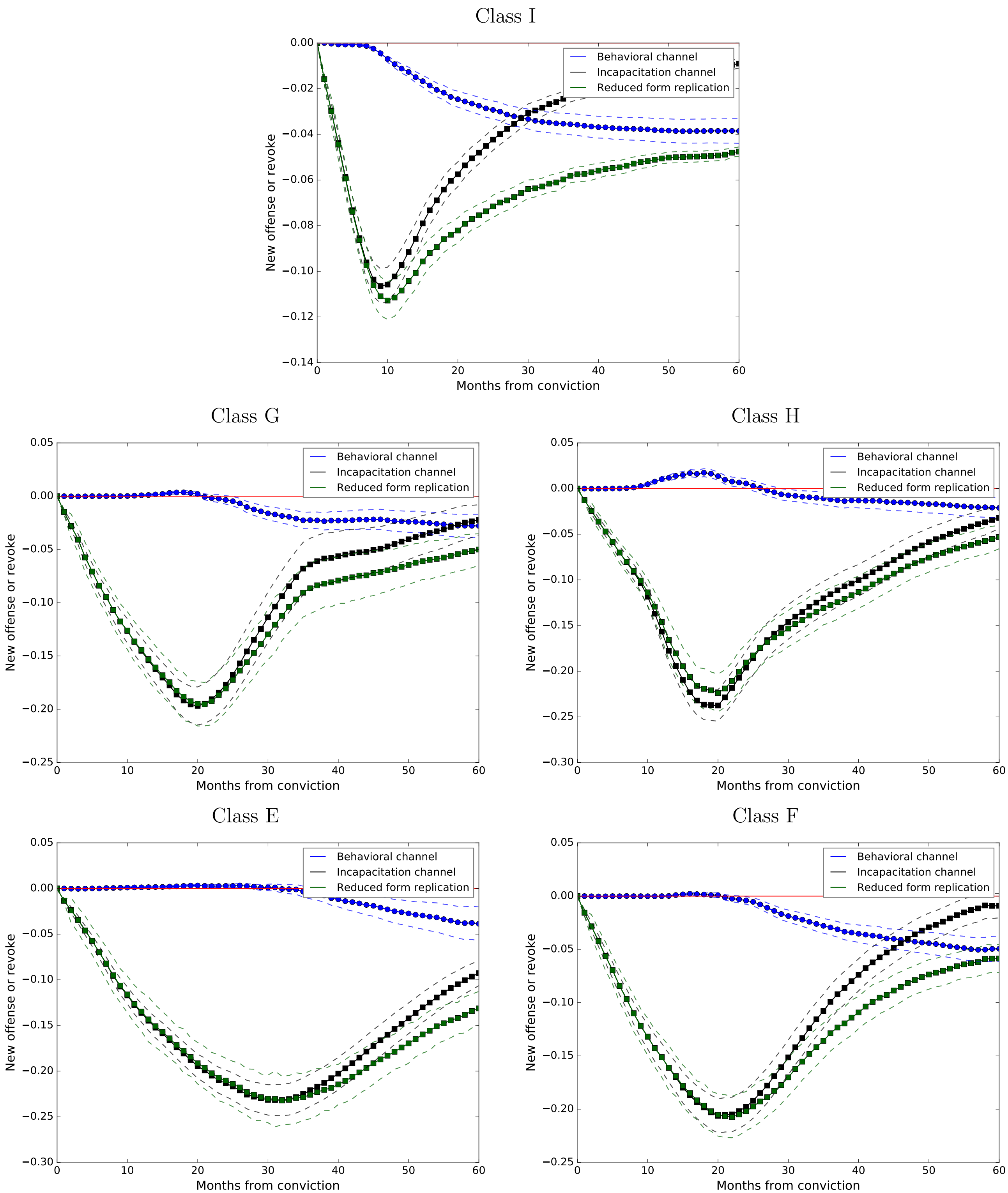


(a) Compliers of class I (least severe offenses)

(b) Compliers of class E (most severe offenses)

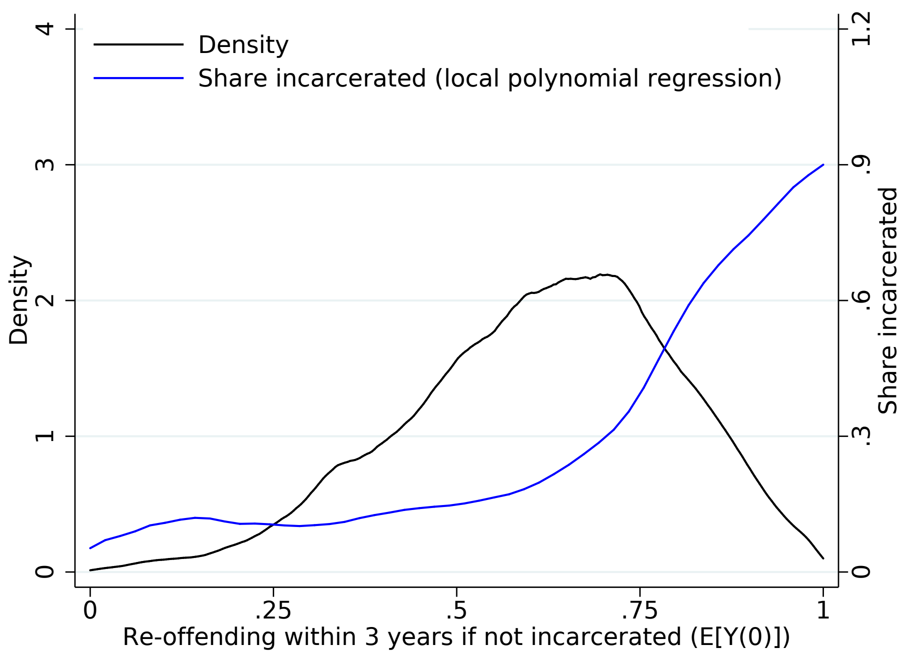
Notes: This figure shows control function estimates of reoffending within t months from release for offenders with characteristics similar to those of compliers in class E (most severe offenses) and class I (least severe offenses). We use the average unobserved heterogeneity of compliers in each offense class, i.e., $\sum_{d=1}^{\bar{D}} E[\nu_i|D_i(1) \geq d > D_i(0)] \cdot \omega_d$. This term is the average $\hat{\nu}$ of compliers in a given offense class. Notice that the weights ω_d are always positive and sum to one since $\omega_d \equiv \frac{\Pr(D_i(1)\geq d>D_i(0))}{\sum_{j=1}^{\bar{D}} \Pr(D_i(1)\geq j>D_i(0))}$; see the description in Section 2 for more details.

Figure 9: Decomposition of reduced form RD estimates by offense class



Notes: This figure shows the results of using the control function estimates to replicate and decompose the reduced form effects on reoffending within t months from conviction into the incapacitation (black line) and behavioral (blue line) channels. The decomposition is preformed as follows. First, we use the model estimates to replicate the reduced form RD estimates (green line). Next we assume that there are no behavioral effects by setting the coefficients on all incarceration variables to zero and replicate the reduced form estimates under this null (black line). The difference between the green and black lines can be attributed to the behavioral channel. We name this unexplained component the "behavioral residual". We calculate standard errors using a block bootstrap procedure with 500 iterations at the individual level to account for within-individual serial correlation.

Figure 10: Distribution of reoffending probabilities and the share of offenders incarcerated



Notes: This figure shows the distribution of reoffending probabilities predicted using the control function estimates and the share of individuals who are incarcerated at each level of predicted risk of reoffending. The x-axis show the predicted likelihood of reoffending, which is measured as committing a new offense or a probation revocation within 3 years of conviction if not incarcerated. The black line (left y-axis) shows the density of each of the predicted reoffending likelihoods. The blue line (right y-axis) shows the share of of individuals actually incarcerated by the current sentencing regime as a function of the reoffending likelihood (x-axis).

Tables

Table 1: Summary Statistics: Demographics, sentencing and reoffending

	Mean (1)	Median (2)	Std. (3)
Demographics:			
Male	0.82	-	0.39
Race			
White	0.43	-	0.49
Black	0.50	-	0.5
Other	0.07	-	0.26
Born in NC	0.65	-	0.48
Age at offense	29.97	28.00	10.20
Age at conviction	30.96	28.75	10.31
Incarceration measures:			
Sentenced to any incarceration	0.32	-	-
Incarceration sentence (months)	4.40	0.00	9.24
Months served (months)	6.44	0.00	15.22
Incarceration sentence conditional on positive sentence (months)	13.55	10.00	11.96
Months served conditional on positive sentence (months)	20.17	14.11	21.19
Recidivism measures from conviction:			
Recidivate in 1 years	0.17	-	-
Felony recidivate in 1 years	0.10	-	-
Recidivate in 2 years	0.29	-	-
Felony recidivate in 2 years	0.19	-	-
Recidivate in 3 years	0.37	-	-
Felony recidivate in 3 years	0.25	-	-
Recidivate in 5 years	0.46	-	-
Felony recidivate in 5 years	0.32	-	-
Recidivate in period	0.55	-	-
Felony recidivate in period	0.41	-	-
Days to recidivate from conviction conditional on recidivating	1006.32	684.00	1035.28
Recidivism measures from at risk:			
Recidivate in 1 years from at risk	0.22	-	-
Felony recidivate in 1 years from at risk	0.14	-	-
Recidivate in 2 years from at risk	0.33	-	-
Felony recidivate in 2 years from at risk	0.22	-	-
Recidivate in 3 years from at risk	0.40	-	-
Felony recidivate in 3 years from at risk	0.27	-	-
Recidivate in 5 years from at risk	0.48	-	-
Felony recidivate in 5 years from at risk	0.34	-	-
Days to recidivate from release conditional on recidivating	878.75	524.00	1008.33
Total N	519,057		
Total unique individuals	322,320		

Notes: This table shows summary statistics for the primary analysis sample. Note that not all observations are included in all regressions, since regressions of outcomes over a fixed horizon (e.g., offending within three years of conviction) are restricted to observations observed over that horizon, cutting off some later dates in the sample. Notice conditional on receiving a positive incarceration sentence offenders spend 20.17 months incapacitated, which is approximately 50% longer than the average sentence length. The difference between average sentences and average incapacitation spells reflects both the fact that sentences represent minimum sentences and 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.

Table 2: Effect of months incarceration on committing any new offense within 3 years of sentencing

	(1) OLS	(2) OLS	(3) RD	(4) RD
Months incar	-0.00458***	-0.00537***	-0.00718***	-0.00794***
	(0.0000411)	(0.0000501)	(0.000766)	(0.000751)
N	491135	491135	491135	491135
Dep. var. mean non-incar.	0.425	0.425	0.425	0.425
Effect of 1 year incar. (pct)	-12.9	-15.2	-20.3	-22.4
Controls	No	Yes	No	Yes
F (excluded-instruments)			205.7	209.7
J stat			3.393	2.663
J stat p			0.494	0.616
Hausman p			0.00785	0.000890

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data between 0 and 3 years of the individual's sentencing date. Standard errors (in parentheses) are clustered by individual. The OLS estimates in Columns 1 and 2 are from estimating Equation (5) using OLS. The 2SLS estimates in columns 3 and 4 are from estimating Equation (4) using 2SLS. The J stat refers to the Sargan-Hansen test of over-identifying restrictions. This test examines the null hypothesis that incarceration has the same effects when estimated using the different instruments under the assumption that the effects are linear and not heterogeneous. Since we have 5 instruments there are five degrees of freedom. The Hausman test examines the null hypothesis that incarceration length assignment is not endogenous by comparing estimates using OLS and using 2SLS under the assumption of linear effects without heterogeneity across individuals. Due to clustering, the F statistic reported is cluster-robust. Effective and non-robust F statistics are similar. The number of observations is smaller than in Table 1 because the sample in the regressions is restricted to individuals that are observed at least three years after the date of sentencing.

Table 3: Effect of months of incarceration on alternative reoffending outcomes within 3 years of sentencing

	Measure of crime					
	(1) Re-incarceration	(2) Any new offense	(3) Felony	(4) Assault	(5) Property	(6) Drug
Months incar	-0.0138***	-0.00794***	-0.00601***	-0.00216***	-0.00327***	-0.00281***
	(0.000724)	(0.000751)	(0.000702)	(0.000435)	(0.000541)	(0.000493)
N	491135	491135	491135	491135	491135	491135
Dep. var. mean among non-incarcerated	0.462	0.425	0.306	0.0690	0.164	0.166
One year effect in percentages	-36.0	-22.4	-23.6	-37.6	-23.9	-20.3
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	209.7	209.7	209.7	209.7	209.7	209.7
J stat	58.25	2.663	6.095	2.505	3.876	11.48
J stat p	6.75e-12	0.616	0.192	0.644	0.423	0.0216
Hausman p	2.99e-12	0.000890	0.00260	0.00588	0.0997	0.256
Lochner-Moretti p	0.00384	0.0173	0.00451	0.0312	0.0501	0.383

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data between 0 and 3 years of the individual's sentencing date. Standard errors (in parentheses) are clustered by individual. Each column represents a different type of new offense. For example, the estimates in Column 2 are the same as the estimates in Column 4 of Table 2. The J stat refers to the Sargan-Hansen test of over-identifying restrictions. This test examines the null hypothesis that incarceration has the same effects when estimated using the different instruments under the assumption that the effects are linear and not heterogeneous. Since we have 5 instruments there are five degrees of freedom. The Hausman test examines the null hypothesis that incarceration length assignment is not endogenous by comparing estimates using OLS and using 2SLS under the assumption of linear effects without heterogeneity across individuals. Due to clustering, the F statistic reported is cluster-robust. Effective and non-robust F statistics are similar. The number of observations is smaller than in Table 1 because the sample in the regressions is restricted to individuals that are observed at least three years after the date of sentencing. The Lochner-Moretti p-values are a generalization of the standard Hausman test of endogeneity to an ordered treatment with multiple levels.

Table 4: Effect of months of incarceration on various reoffending outcomes within 3 years of at-risk

	Measure of crime					
	(1)	(2)	(3)	(4)	(5)	(6)
	Re-incarceration	Any new offense	Felony	Assault	Property	Drug
Months incar	-0.00745***	-0.000383	0.000566	-0.000320	0.00113	0.00118
	(0.000887)	(0.000885)	(0.000871)	(0.000620)	(0.000736)	(0.000695)
N	477689	477689	477689	477689	477689	477689
Dep. var. mean among non-incarcerated	0.462	0.425	0.305	0.0690	0.164	0.166
One year effect in percentages	-19.3	-1.08	2.22	-5.57	8.29	8.57
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	257.6	257.6	257.6	257.6	257.6	257.6
J stat	47.84	5.931	4.053	1.329	8.411	8.381
J stat p	1.02e-09	0.204	0.399	0.856	0.0776	0.0786
Hausman p	0.171	0.704	0.899	0.561	0.814	0.0809
Lochner-Moretti stat	-0.000710	-0.000554	-0.000933	-0.000563	-0.000209	0.000822
Lochner-Moretti p	0.426	0.537	0.288	0.362	0.776	0.240

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data between 0 and 3 years of the individual's release date. Standard errors are clustered by individual. Each column represents a different type of new offense (e.g., drug, property). The J stat refers to the Sargan-Hansen test of over-identifying restrictions. This test examines the null hypothesis that incarceration has the same effects when estimated using the different instruments under the assumption that the effects are linear and not heterogeneous. Since we have 5 instruments there are five degrees of freedom. The Hausman test examines the null hypothesis that incarceration length assignment is not endogenous by comparing estimates using OLS and using 2SLS under the assumption of linear effects without heterogeneity across individuals. Due to clustering, the F statistic reported is cluster-robust. Effective and non-robust F statistics are similar. The Lochner-Moretti p-values are a generalization of the standard Hausman test of endogeneity to an ordered treatment with multiple levels. The number of observations is smaller than in Table 1 because the sample in the regressions is restricted to individuals that are observed at least three years from the date of being at-risk.

Table 5: Control Function Estimates: Any new offense or probation revocation within 3 years of at-risk

	OLS	2SLS	CF		
	(1)	(2)	(3)	(4)	(5)
Any incarceration	-0.0101**	-0.204	-0.0308***	-0.0635***	-0.0638***
	(0.00345)	(0.133)	(0.00476)	(0.00710)	(0.00711)
Years incar	-0.0611***	-0.0128	-0.0705***	-0.0832***	-0.0837***
	(0.00400)	(0.175)	(0.00427)	(0.00684)	(0.00686)
Years incar square	0.00862***	0.00916	0.00864***	0.0163***	0.0166***
	(0.000834)	(0.0438)	(0.000834)	(0.00163)	(0.00168)
$\hat{\nu}$ (selection on unobserved criminality)			0.0196***	0.0634***	0.0659***
			(0.00315)	(0.00672)	(0.00731)
Any incarceration × $\hat{\nu}$				-0.0191**	-0.0229**
				(0.00640)	(0.00765)
Years incar × $\hat{\nu}$				-0.0121*	-0.0125*
				(0.00511)	(0.00515)
Years incar square × $\hat{\nu}$				-0.00118	-0.00128
				(0.00104)	(0.00105)
$\hat{\nu}^2$					0.00145
					(0.00167)
Marginal effects of years of incarceration (in %)					
1 year incarceration effect (%)	-10.99	-28.86	-16.28	-22.91	-22.99
SE	0.313	5.200	0.891	1.201	1.204
3 year incarceration effect (%)	-20.36	-29.14	-28.90	-29.29	-29.06
SE	0.563	6.078	1.453	2.020	2.040
2 to 3 years incarceration effect (%)	-3.167	8.233	-4.793	-0.334	-0.123
SE	0.245	4.877	0.349	0.710	0.755
Obs.	477616	477616	477616	477616	477616
Dep. mean of non-incarcerated	0.569	0.569	0.569	0.569	0.569
Age at release FEs	Yes	Yes	Yes	Yes	Yes
Year of release FEs	Yes	Yes	Yes	Yes	Yes
J-stat (punishment type discontinuities)		4.880	14.85	2.986	
J-stat p (punishment type discontinuities)		0.0872	0.00503	0.0840	
J-stat (all discontinuities)		44.42	54.27	34.12	29.11
J-stat p (all discontinuities)		0.000296	0.0000301	0.00524	0.0156

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows estimates of several specifications of the control function approach and a comparison of the estimates to 2SLS and OLS estimates. The dependent variable is an indicator for any charges or probation revocation recorded in the AOC or DPS data between 0 and 3 years of the individual's release date. Standard errors are clustered by individual. The estimated control function $\widehat{\lambda}\left(X_i, Z_i^l, d\right) = \mathbb{E}\left[\nu_i | X_i, Z_i^l, D_i = d\right]$ is denoted by $\hat{\nu}$ in the table and $\hat{\nu}^2 = \mathbb{E}\left[\nu^2 | X_i, Z_i^l, D_i = d\right]$. The marginal effects show the impacts of exposure to incarceration normalized by the rate of reoffending among non-incarcerated individuals. The J-tests at the bottom of the table show model fit diagnostics. The J-test for the control function models comes from a 2SLS estimation of each specification when the endogenous variables are the $\hat{\nu}$ terms and the instruments are as specified in the table. This provides an over-identification test for the coefficients on the control function terms.

Table 6: Heterogeneity in Control Function Estimates: Any new offense or probation revocation within 3 years of at-risk

	Population	Compliers				
	(1) All	(2) Class E	(3) Class F	(4) Class G	(5) Class H	(6) Class I
1 year incarceration effect (%)	-22.91***	-28.53***	-28.24***	-20.45***	-20.87***	-28.64***
	(1.201)	(1.437)	(1.443)	(0.991)	(1.059)	(1.893)
3 year incarceration effect (%)	-29.29***	-35.43***	-35.47***	-24.10***	-26.07***	-40.83***
	(2.020)	(2.591)	(2.534)	(1.992)	(1.883)	(2.490)
2 to 3 years incarceration effect (%)	-0.334	0.323	0.0438	1.159	0.126	-3.419***
	(0.710)	(0.998)	(0.945)	(0.861)	(0.713)	(0.466)
$\sum_{d=1}^{D} E[\nu_i \mid D_i(1) \geq d > D_i(0)] \cdot \Pr(D_i(1) \geq d > D_i(0))$ (average ν)	0	-0.184	-0.117	-0.470	-0.148	0.970
$\sum_{d=1}^{\bar{D}} d \cdot \Pr(D_i(1) \geq d > D_i(0))$ (average incarceration exposure)	.	21.57	15.23	15.30	19.71	7.586
Dep. mean of non-incarcerated	0.569	0.436	0.448	0.563	0.602	0.565
Age at release FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year of release FEs	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows estimates of the marginal effects of incarceration by felony class. The dependent variable is an indicator for any charges or probation revocation recorded in the AOC or DPS data between 0 and 3 years of the individual's release date. All estimates are based on the same specification as in Column 4 of Table 5. All the marginal treatment effects are expressed in % terms relative to the mean reoffending rate among non-incarcerated offenders. Column 1 reports population treatment effects, i.e., when $\hat{\nu} = 0$. Columns 2-5 report estimates using the average unobserved heterogeneity ($\hat{\nu}$) of compliers in each offense class, i.e., $\sum_{d=1}^{\bar{D}} E[\nu_i | D_i(1) \geq d > D_i(0)] \cdot \omega_d$. This term is the average $\hat{\nu}$ of compliers in a given offense class. Notice that the weights ω_d are always positive and sum to one since $\omega_d \equiv \frac{\Pr(D_i(1) \geq d > D_i(0))}{\sum_{j=1}^{\bar{D}} \Pr(D_i(1) \geq j > D_i(0))}$; see the description in Section 2 for more details. Similarly $\sum_{d=1}^{\bar{D}} d \cdot \omega_d$ is the average change in exposure to incarceration due to a punishment type discontinuity expressed in terms of months of incarceration. Standard errors are clustered by individual.

Table 7: Share of reduced form RD estimates attributable to behavioral channel

	Class E (1)	Class F (2)	Class G (3)	Class H (4)	Class I (5)
One year	0.0095	0.00044	0.0061	0.062	0.1
	(0.0034)	(0.0011)	(0.0024)	(0.0074)	(0.0083)
Three years	0.0221	0.229	0.26	0.0929	0.612
	(0.0133)	(0.0251)	(0.0546)	(0.0258)	(0.0198)
Five years	0.295	0.845	0.556	0.398	0.81
	(0.0458)	(0.0714)	(0.0902)	(0.0678)	(0.017)

Notes: This table shows the results of decomposing the model based replications of the reduced form RD estimates into the behavioral and incapacitation channels. Each cell shows the share of the reduced form estimates that is explained by the behavioral channel. The outcome is any new offense or probation revocation within 1, 3 or 5 years from the date of conviction, as indicated in each row. Standard errors are calculated using a block bootstrap procedure at the individual level with 500 iterations.

Table 8: Break-even Estimates: Dollar values of social cost of crime necessary to justify the costs of incarceration

	(1) All	(2) Class E	(3) Class F	(4) Class G	(5) Class H	(6) Class I
			8 year from sentencing			
New offense	91784**	89129*	201218	51469	104583	59030
	(30349)	(37893)	(277991)	(37672)	(121487)	(55113)
	[-0.0298]	[-0.0307]	[-0.0136]	[-0.0532]	[-0.0262]	[-0.0464]
New offense or probation revoke	41904***	45768***	63896	28236*	44061	22935*
	(7787)	(12306)	(33759)	(14325)	(25265)	(11526)
	[-0.0653]	[-0.0598]	[-0.0429]	[-0.0970]	[-0.0621]	[-0.1194]
New felony offense	164081*	170573	-681509	174133	122755	45404
	(75868)	(102653)	(2460344)	(310123)	(141428)	(29223)
	[-0.0167]	[-0.0161]	[0.0040]	[-0.0157]	[-0.0223]	[-0.0603]
			3 year from sentencing			
New offense	77142***	127966***	85723***	78757*	56041***	23584**
	(9095)	(27123)	(25746)	(30634)	(13718)	(8002)
	[-0.0355]	[-0.0214]	[-0.0319]	[-0.0348]	[-0.0489]	[-0.1161]
New offense or probation revoke	39145***	60748***	41342***	36908***	32322***	10621***
	(2789)	(7304)	(7157)	(7667)	(5184)	(2380)
	[-0.0699]	[-0.0451]	[-0.0662]	[-0.0742]	[-0.0847]	[-0.2578]
New felony offense	118374***	262816**	155565*	120425*	69214***	30754**
	(18109)	(96199)	(72854)	(58059)	(18071)	(11393)
	[-0.0231]	[-0.0104]	[-0.0176]	[-0.0227]	[-0.0396]	[-0.0890]
			1 year from sentencing			
New offense	58809***	93314***	76821***	73560***	38258***	36417***
	(3837)	(12699)	(10716)	(10781)	(4081)	(7458)
	[-0.0466]	[-0.0293]	[-0.0356]	[-0.0372]	[-0.0716]	[-0.0752]
New offense or probation revoke	35021***	52154***	47438***	46549***	26803***	17217***
	(1664)	(4782)	(5109)	(5177)	(2308)	(2163)
	[-0.0782]	[-0.0525]	[-0.0577]	[-0.0588]	[-0.1022]	[-0.1590]
New felony offense	83042***	151594***	104113***	114349***	51761***	50428***
	(6500)	(27109)	(16702)	(21134)	(6528)	(12262)
	[-0.0330]	[-0.0181]	[-0.0263]	[-0.0239]	[-0.0529]	[-0.0543]

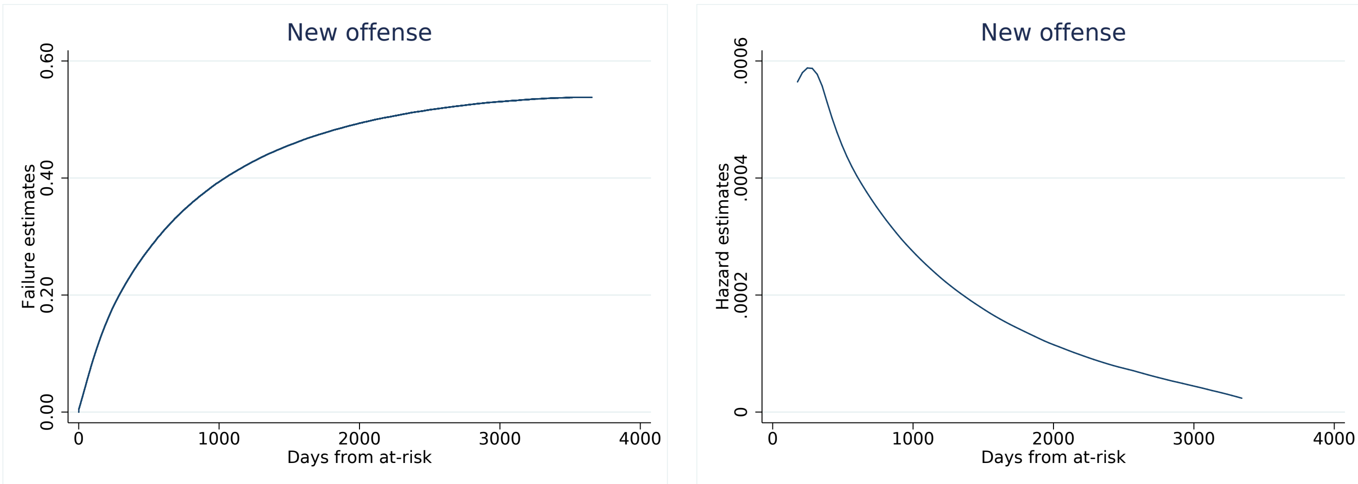
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows estimates of Equation (16), i.e., the break-even, dollar value social cost of crime required to justify the fiscal costs of incarceration. The first coefficient in every cell reports the ratio between a 2SLS coefficient of cumulative incarceration time (in months) from conviction on cumulative new offenses (β_{2SLS}) and the cost of incarcerating an offender for one month, i.e., $\frac{\beta_{2SLS}}{\text{One month incarceration cost}}$. The second value (in parenthesis) reports the standard error of the break-even value. Lastly, the third estimate (in square brackets) reports the 2SLS coefficient β_{2SLS} before we divide it by the average cost of a month of incarceration (relative to probation) according to the North Carolina Department of Public Safety, which is \$2,738.1. Each panel contains estimates for cumulative measures measured over a different time horizon — 1, 3, and 8 years from conviction. Within each panel, each row contains estimates for a different measure of reoffending measure (e.g., any new offense, any new felony offense). Standard errors are clustered at the individual level. The costs of incarceration and probation in North Carolina were taken from the following source: https://www.ncdps.gov/adult-corrections/cost-of-corrections.

A Additional figures and tables

Figure A.1: Failure and hazard functions of committing a new offense since being at-risk to reoffend

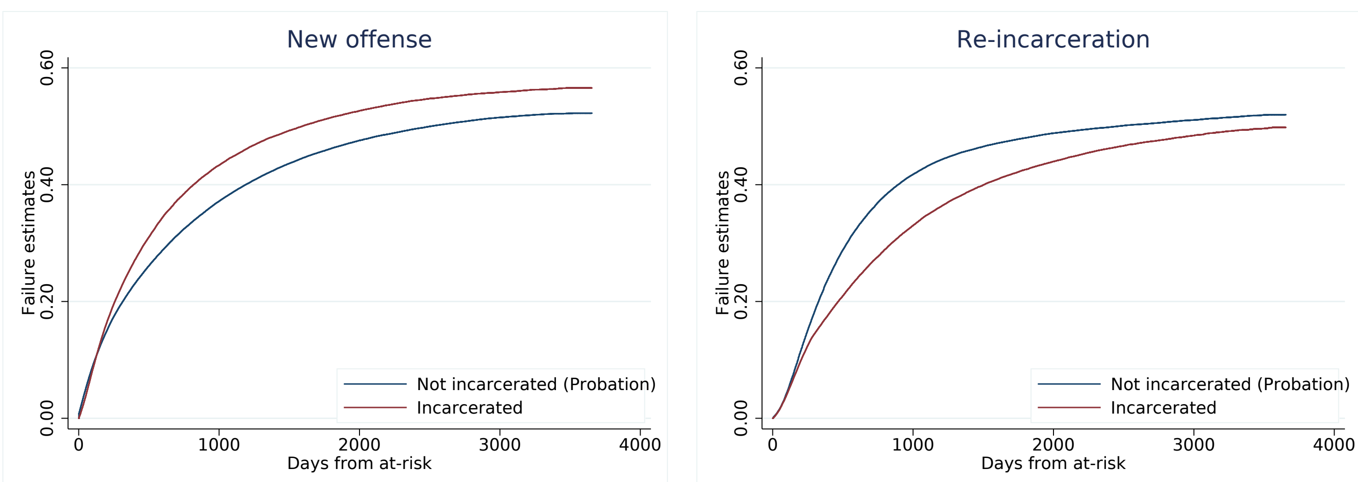


(a) Failure function

(b) Hazard function

Notes: Panel (a) shows the failure function of committing any new offense within t months from being in the community and at-risk to offend. Panel (b) shows the hazard function, i.e., the likelihood of committing an offense at time t conditional on not reoffending prior to that time. For individuals not incarcerated, the at-risk date is the conviction date. For the incarcerated, the at-risk date is the date of release from incarceration.

Figure A.2: Failure functions of initially incarcerated and non-incarcerated offenders, new offenses vs. reincarceration

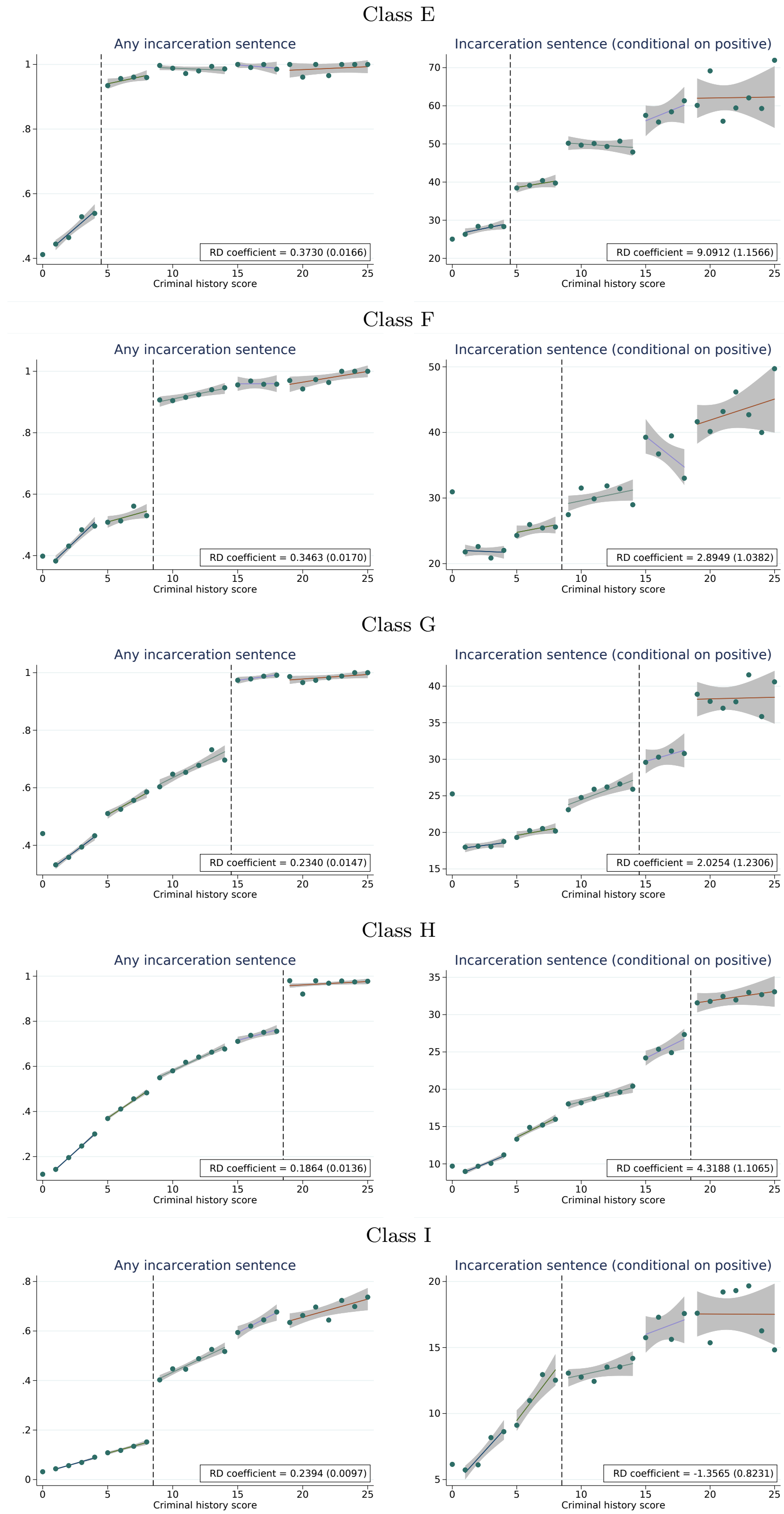


(a) New offense

(b) Reincarceration

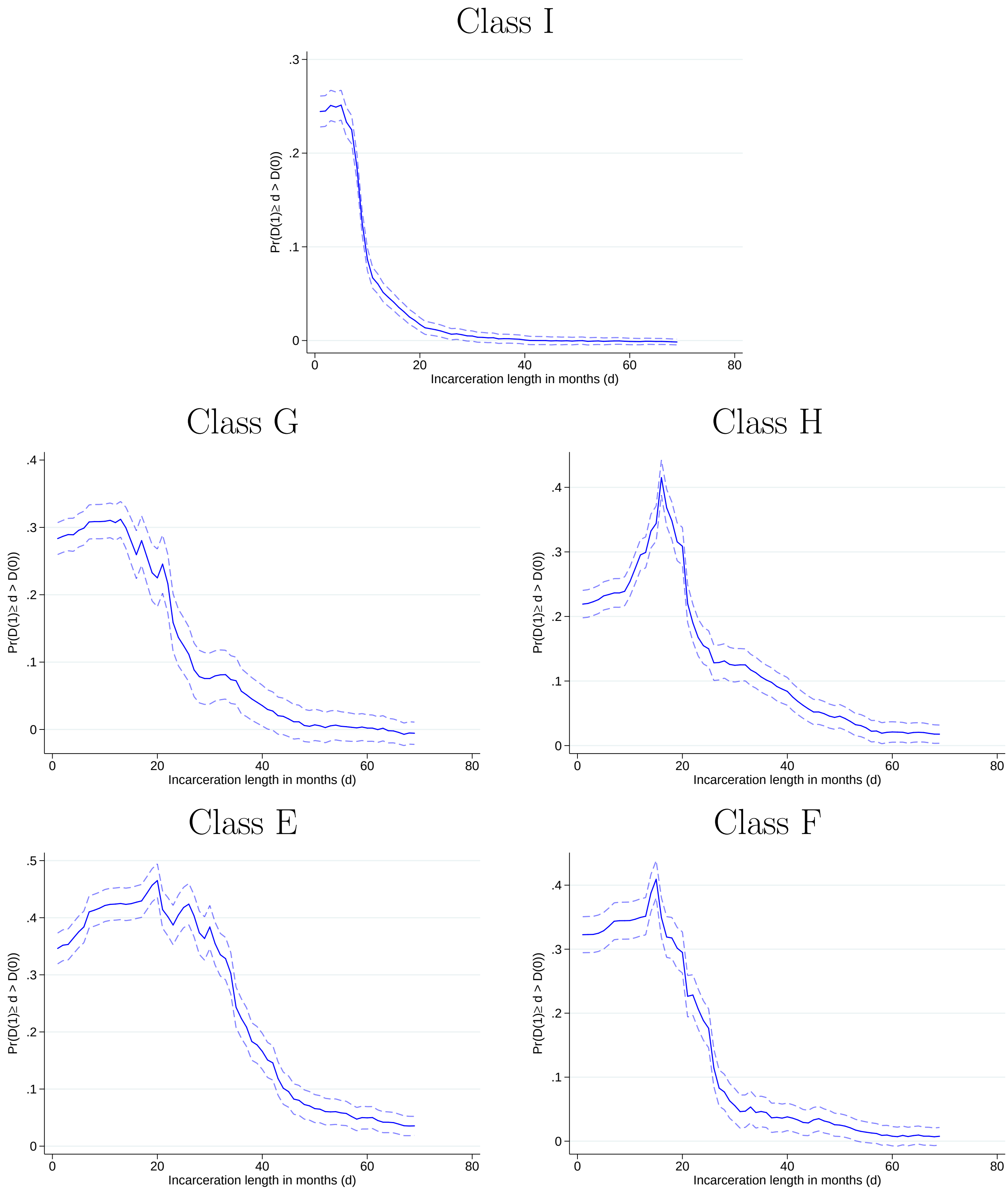
Notes: This figure shows the failure function for either committing a new offense (panel a) or being reincarcerated (panel b). This figure highlights how measurement of reoffending/recidivism can impact differences between incarcerated offenders and those assigned to a probation regime. In Panel (a), the offenders assigned to incarceration commit more crime once they are released; however, when measuring reoffending as a reincarceration event the difference between the two groups of offenders changes signs.

Figure A.3: First stage: Sentencing outcomes by felony class and prior points



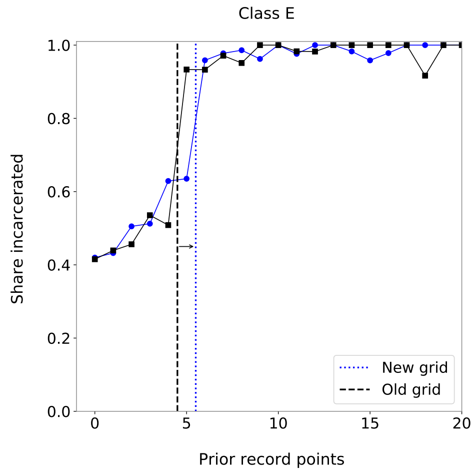
Notes: The x-axis in all plots is the number of prior record points. The y-axis is the share of offenders who are sentenced to incarceration (left plots) or the number of months incarcerated conditional on a positive sentence (right plots). The figures only include offenses sentenced under the sentencing grid that applied to offenses committed between 1995 to 2009. In 2009 the guidelines changed and the discontinuities shifted by one prior points either to the left or to the right. All official grids are in Appendix D.

Figure A.4: Average causal response (ACR) weights across punishment type discontinuities

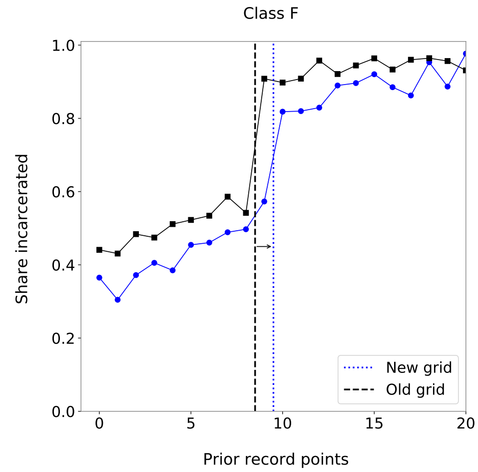


Notes: Each figure plots estimates of the shifts in incarceration exposure generated by each instrument, which correspond to the weights in the average causal response. These shifts reflect the probability an offender would spend less than d months incarcerated if assigned $Z_i = 0$, but at least d months if assigned $Z_i = 1$. This probability can be estimated non-parametrically as $\mathbb{E}\left[1(D_i \geq d)|Z_i = 1\right] - \mathbb{E}\left[1(D_i \geq d)|Z_i = 0\right]$, which corresponds to the coefficient on Z_i in our first stage specification when $1(D_i \geq d)$ is the outcome. Panel (a) shows only offenses sentenced under the sentencing grid that applied to offenses committed between 1995 to 2009. In 2009 the guidelines changed and the discontinuities shifted by one prior points either to the left or to the right. All official grids are in Appendix D. Standard errors are clustered at the individual level.

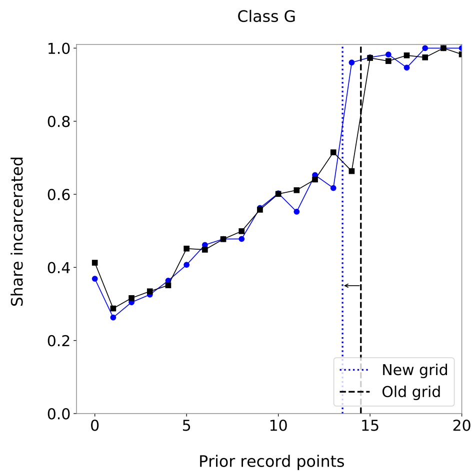
Figure A.5: Shifts in incarceration exposure as a result of 2009 grid changes



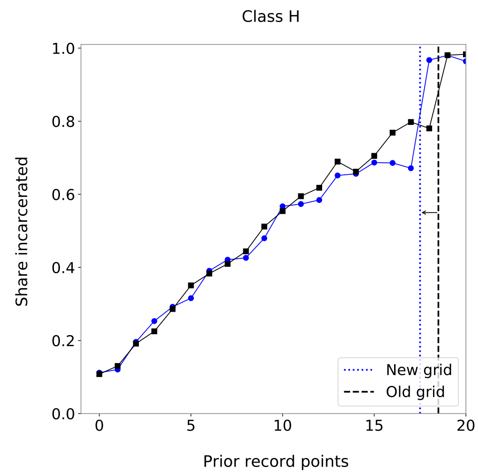
(a) Class E



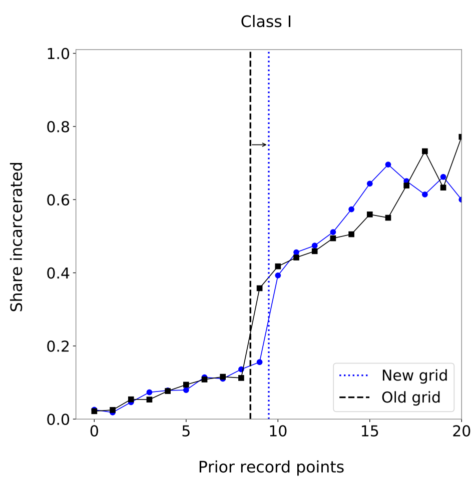
(b) Class F



(c) Class G



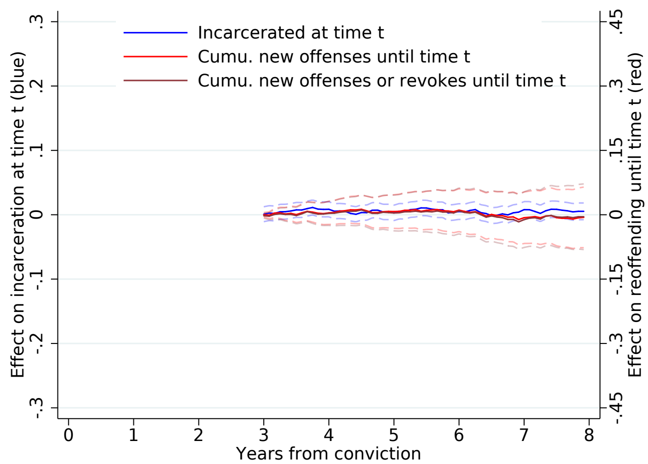
(d) Class H



(e) Class I

Notes: The x-axis in all plots is the number of prior record points. The y-axis is the share of offenders who are sentenced to an incarceration punishment. The black line represents the share of offenders sentenced to incarceration prior to the 2009 reform, with the blue line plotting the share afterwards. The plots demonstrate how the discontinuities in the sentencing grid, and thus exposure to incarceration, changed following the 2009 change in sentencing guidelines. The old grid refers to the sentencing grid in place between 1995 to 2009; the new grid refers to the sentencing in place from 2009 to 2011 (see Appendix D). The location of the discontinuities in the punishment type and severity did not change since the 2009 reform to the present, although sentence lengths within each cell have been adjusted slightly.

Figure A.6: Reduced form estimates of *cumulative* reoffending up to period t from conviction

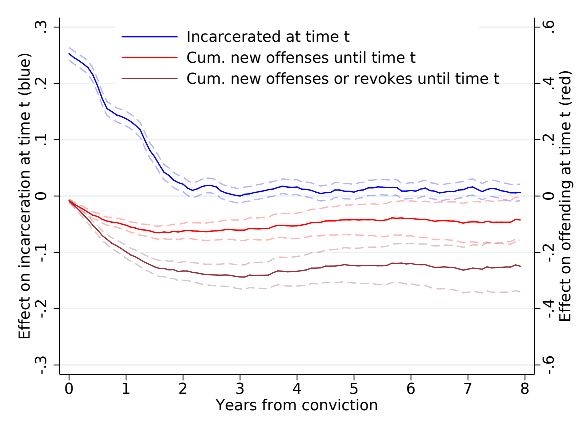


Cumulative reoffending from 3 years on

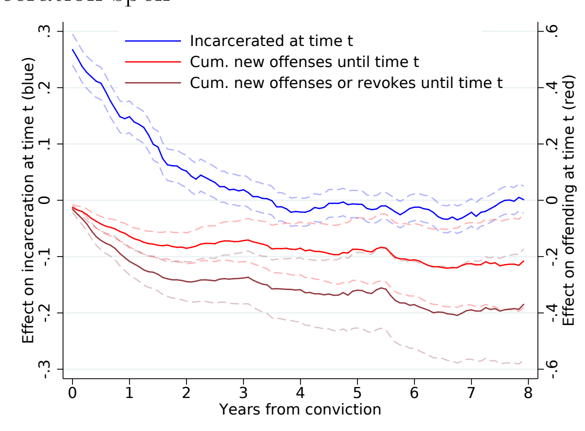
Notes: This figure shows the reduced form effects of being to the right of a punishment type discontinuity on several key outcomes. The blue line (left y-axis) in all panels shows effects on an indicator for being incarcerated at any point *in* month t from conviction. The red line (right y-axis) reports effects on the cumulative number of new offenses committed from month 36 until month t after conviction, with the maroon line including probation revocations. Standard errors are clustered by individual. Each point in each figure is an estimate of γ^{RF} for the relevant outcome for month t. This estimate is a constrained version of Equation 4 that requires the coefficients on all instruments to be the same (i.e., $\gamma^2_{E,4} = \gamma^2_{F,9} = \gamma^2_{G,14} = \gamma^2_{H,19} = \gamma^2_{I,9} = \gamma^{RF}$). This strategy averages across all five offense classes and instruments, but collapses our variation into a single coefficient. γ^{RF} can therefore be thought of as the average reduced form effect across the five punishment type discontinuities (taking the actual average of the individual reduced forms yields highly similar results). The notation used is based on the guidelines in place prior to the 2009 reform, although all observations are used in estimation. The regression specifications include as controls demographics (e.g., race, gender, age FEs), FEs for the duration of time previously incarcerated, the number of past incarceration spells and the number of past convictions, county FEs, and year FEs. Estimates without controls yield similar results (see Table 2).

Figure A.7: Heterogeneity in the reduced form estimates on *cumulative* reoffending by offender characteristics

Previous incarceration spell

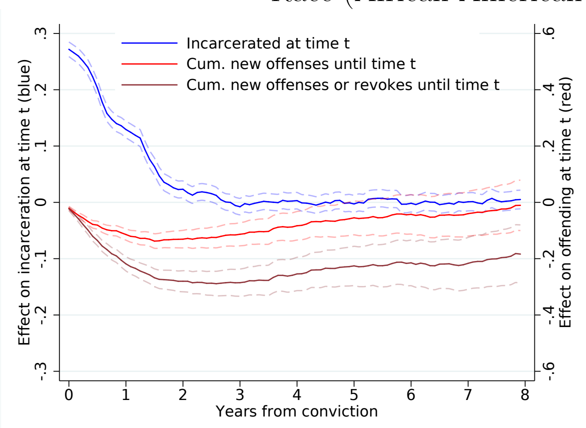


(a) Previously incarcerated

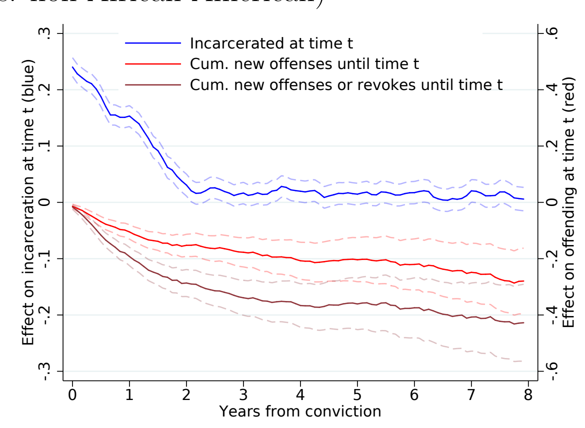


(b) Not previously incarcerated

Race (African American vs. non-African American)

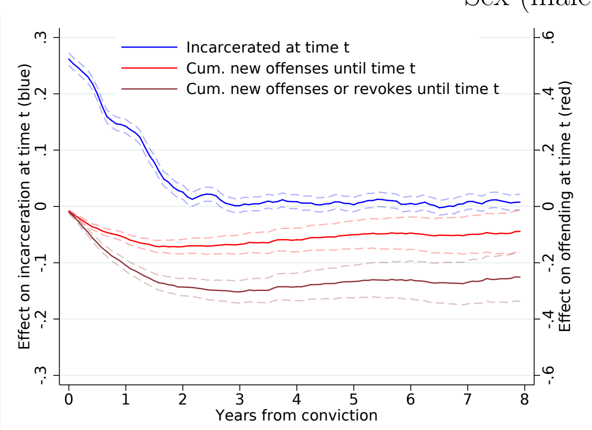


(a) African American

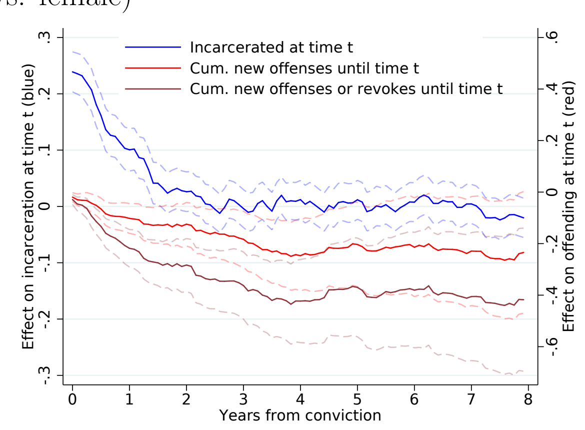


(b) Non-African American

Sex (male vs. female)



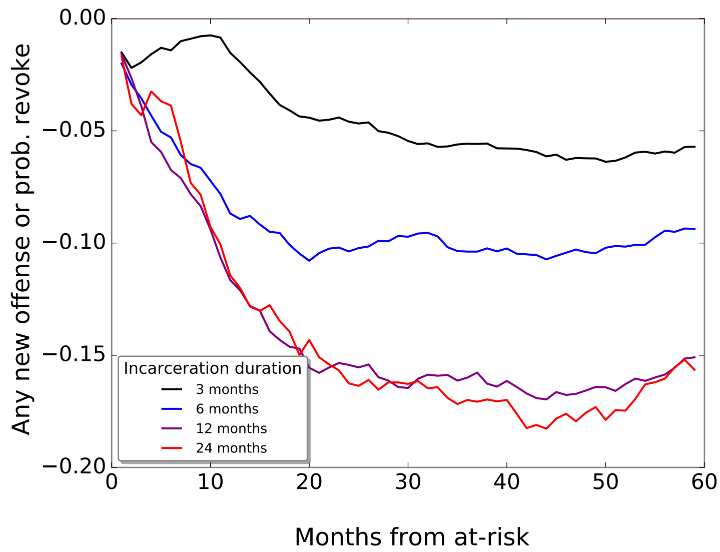
(a) Male



(b) Female

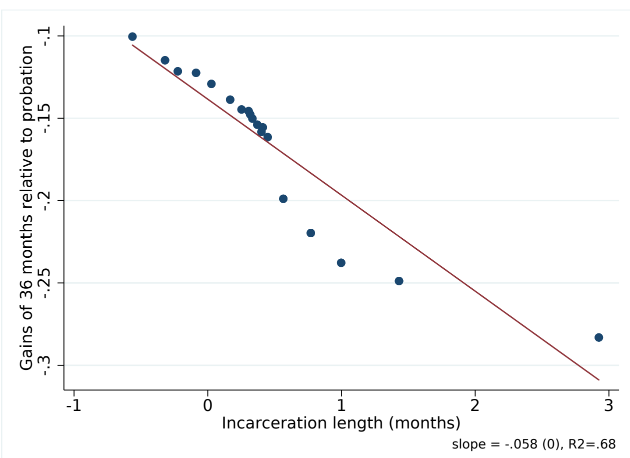
Notes: See the notes in Figure 6.

Figure A.8: Population average treatment effects: Reoffending within t months from at-risk

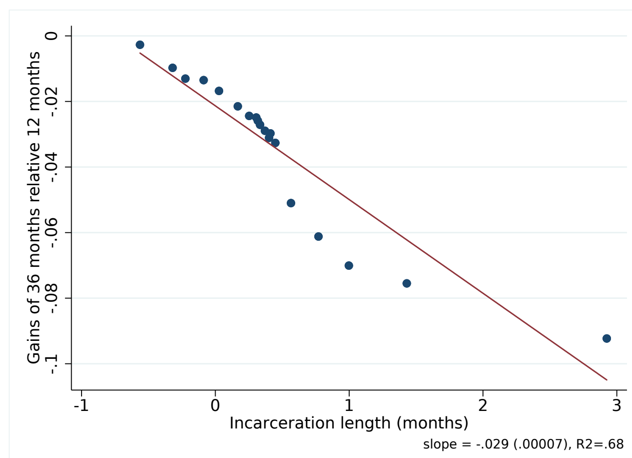


Notes: This figure shows the average treatment effects of different levels of incarceration spells on reoffending with t months after at-risk. The estimates are from the more general selection model described in Equation (11).

Figure A.9: Selection into incarceration based on gains (reduction in reoffending due to exposure)



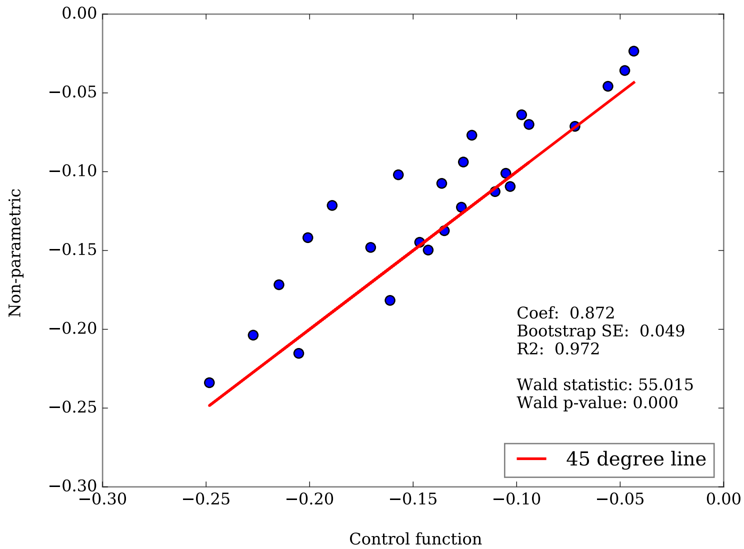
Gains: $D = 36$ relative to $D = 0$ ($\mathbb{E}\left[Y_{i,t+36}(36) - Y_{i,t}(0)\right]$)



Gains: $D = 36$ relative to $D = 12$ ($\mathbb{E}\left[Y_{i,t+36}(36) - Y_{i,t+12}(12)\right]$)

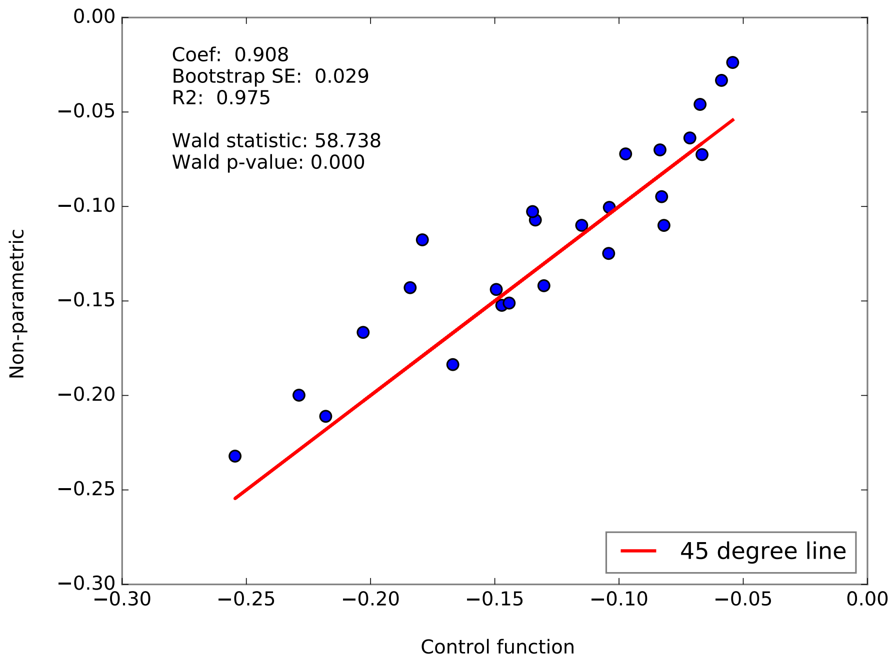
Notes: This figure shows binned scatter plots of the correlation between assignment to longer spells of incarceration and two different types of measures of the gains to incarceration, as denoted in the subtitles.

Figure A.10: Restricted control function goodness of fit: Replication of reduced form RD estimates of reoffending within a time window from conviction



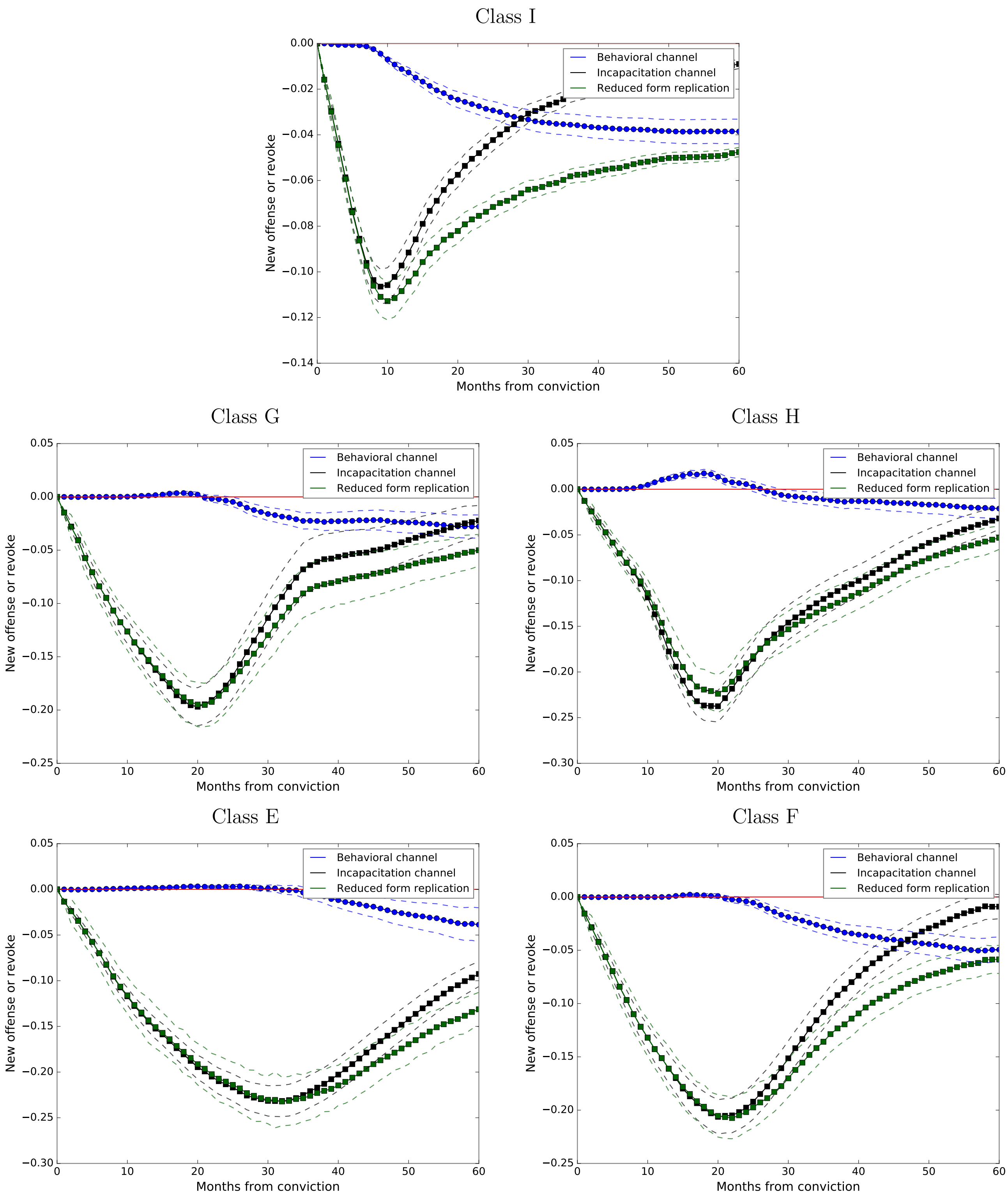
Notes: This figure tests whether the control function estimates the simplified model — which uses a polynomial in D_i and an indicator for any incarceration sentence—can reproduce the quasi-experimental variation induced by the five primary instruments. The y-axis shows the non-parametric RD estimates of the effect of being to the right of a discontinuity ($\mathbb{E}[Y_{i,t}|Z_i=1] - \mathbb{E}[Y_{i,t}|Z_i=0]$) on reoffending for each of the five felony classes for five time horizons (1, 2, 3, 4, and 5 years from *conviction*), generating a total of 25 points. Each estimate is plotted against the control function replication of the same parameter (x-axis). As was shown by Angrist and Imbens (1995) the reduced form of a treatment with multiple levels can be expressed as $\mathbb{E}[Y_{i,t}|Z_i=1] - \mathbb{E}[Y_{i,t}|Z_i=0] = \sum_{d=1}^{\bar{D}} \mathbb{E}[Y_{i,t}(d) - Y_{i,t}(d-1)|D_i(1) \geq d > D_i(0)] \Pr(D_i(1) \geq d > D_i(0))$. Using the ordered choice model, we first replicate the weights / complier probabilities ($Pr(D_i(1) \geq d > D_i(0))$). As shown in Figure M.2, the selection model captures the changes in incarceration exposure induced by the instruments well. Next, we replicate each treatment effects $\mathbb{E}[Y_{i,t}(d) - Y_{i,t}(d-1)|D_i(1) \geq d > D_i(0)]$ using the control function estimates of the model parameters. The combination of weights and treatment effect estimate allows us to construct the control function replication of each reduced form. The red line shows the 45 degree line. If the control function approach perfectly replicated the reduced forms, all the points would lie on the 45 degree line. The Wald statistic and p-value is for a joint test that all the points are one the red line (Coef=1 and R^2 =1). A comparison of reduced form estimates from sentencing to the model-based replications jointly tests the goodness of fit of the selection model described by the ordered-choice model and the parametric restrictions imposed on $\mathbb{E}[Y_{i,t}(d)]$ by the control function approach. Note that we do not include any time-varying controls in the control function specification when we use it replicate the reduced form effects, since the reduced form estimates do not include adjustments for time-varying factors as is discussed in the Section 2.

Figure A.11: Full control function goodness of fit: Replication of reduced form RD estimates of reoffending within a time window from conviction



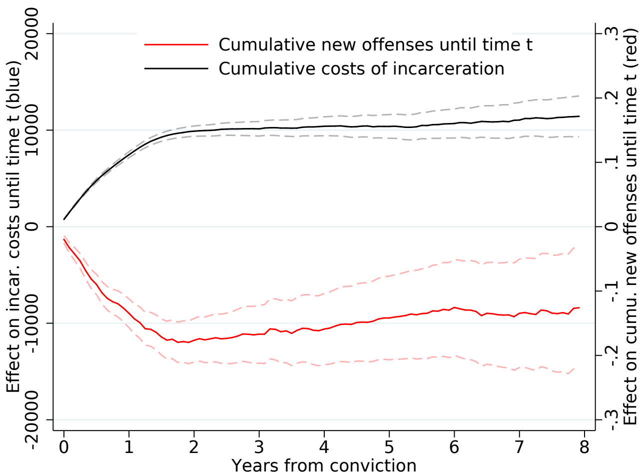
Notes: This figure tests whether the control function estimates the full model in Equation (11) can reproduce the quasi-experimental variation induced by the five primary instruments. The y-axis shows the non-parametric RD estimates of the effect of being to the right of a discontinuity ($\mathbb{E}[Y_{i,t}|Z_i = 1] - \mathbb{E}[Y_{i,t}|Z_i = 0]$) on reoffending for each of the five felony classes for five time horizons (1, 2, 3, 4, and 5 years from *conviction*), generating a total of 25 points. Each estimate is plotted against the control function replication of the same parameter (x-axis). As was shown by Angrist and Imbens (1995) the reduced form of a treatment with multiple levels can be expressed as $\mathbb{E}[Y_{i,t}|Z_i = 1] - \mathbb{E}[Y_{i,t}|Z_i = 0] = \sum_{d=1}^{\bar{D}} \mathbb{E}[Y_{i,t}(d) - Y_{i,t}(d-1)|D_i(1) \geq d > D_i(0)] \Pr(D_i(1) \geq d > D_i(0))$. Using the ordered choice model, we first replicate the weights / complier probabilities ($Pr(D_i(1) \geq d > D_i(0))$). As shown in Figure M.2, the selection model captures the changes in incarceration exposure induced by the instruments well. Next, we replicate each treatment effects $\mathbb{E}[Y_{i,t}(d) - Y_{i,t}(d-1)|D_i(1) \geq d > D_i(0)]$ using the control function estimates of the model parameters. The combination of weights and treatment effect estimate allows us to construct the control function replication of each reduced form. The red line shows the 45 degree line. If the control function approach perfectly replicated the reduced forms, all the points would lie on the 45 degree line. The Wald statistic and p-value is for a joint test that all the points are one the red line (Coef=1 and R^2 =1). A comparison of reduced form estimates from sentencing to the model-based replications jointly tests the goodness of fit of the selection model described by the ordered-choice model and the parametric restrictions imposed on $\mathbb{E}[Y_{i,t}(d)]$ by the control function approach. Note that we do not include any time-varying controls in the control function specification when we use it replicate the reduced form effects, since the reduced form estimates do not include adjustments for time-varying factors as is discussed in the Section 2.

Figure A.12: Decomposition, by offense class, of control function based replications of the reduced form RD estimates to incapacitation and behavioral channels

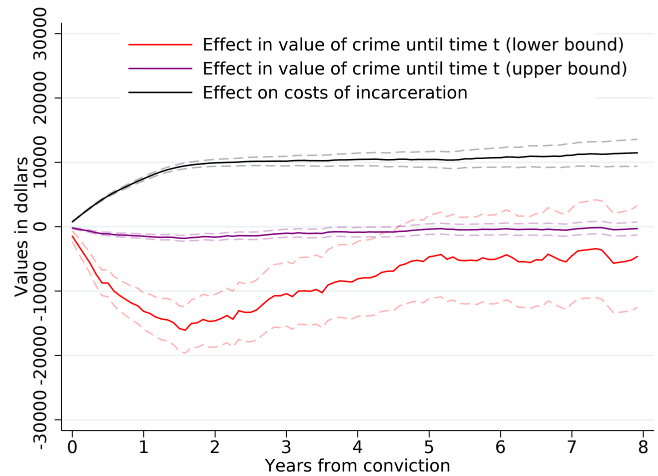


Notes: This figure shows the results of using the control function estimates to replicate and decompose the reduced form RD estimates of reoffending within t months from conviction. The decomposition of the estimates to the incapacitation (black line) and behavioral (blue line) channels is done using the null of no behavioral effects. We first use the CF estimates to replicate the reduced form RD estimates (green line). Next we assume that there are no behavioral effects, i.e., we impose that the coefficients on all the incarceration variables/indicators are equal to zero, and replicate the RD estimates under this null (black line). The difference between the green and black lines is the unexplained part (blue line) in the estimates of the reduced forms and it can be attributed to the behavioral channel. We name this unexplained component the "behavioral residual". We calculate SEs using a block bootstrap procedure with 500 iterations at the individual level to account for within-individual serial correlation.

Figure A.13: Reduced form effects on cumulative number of new offense (crime averated) and cumulative costs of incarceration



(a) Num. new offenses vs. correctional costs



(b) Cumulative value of crime vs. correctional costs

Notes: This figure shows reduced form estimates of being to the right of a punishment type discontinuity on several cumulative measures. The red line in Panel (a) shows effects on the cumulative number of new offenses committed *up to* month t from conviction. The black line (in both panels) is the effect on cumulative costs of incarceration, which is the cumulative months incarcerated up to period t multiplied by the average additional costs incurred by incarcerating an offender for a month instead of placing them on probation. The red and purple lines in Panel (b) show effects on the cumulative dollar value of crime averted. The value of crime averted is calculated by multiplying each criminal event with the appropriate dollar value reported in Appendix Table A.10. The table includes upper and lower bounds; effects on the value of crime are calculated for each of these bounds. All outcomes/measures are with respect to the conviction date. The red line (right y-axis) reports the reduced form effect on the upper bound of the cumulative value of crime averted and the purple line the lower bound. For the estimation details see the main text or the notes in Figure 6.

Table A.1: Average age of study population in other criminal justice studies in the U.S.

Study	Topic	population	State	Ave. age
Muller-Smith (2015)	The effects of incarceration on recidivism	Felony defendants	Texas	**30.26**
Dobbie, Goldin, and Yang (2018)	The Effects of Pretrial Detention on Conviction, Future Crime, and Employment	Age at bail decision	Florida	**33**
Butcher, Park and Piehl (2017)	Disparities in punishment	Convicted	Kansas	**30.926**
Mueller-Smith and Schnepel (2017)	Diversion in the Criminal Justice System	Convicted	Texas	**30.23**
Abrams and Ryan Fackler (2017)	Plea bargaining	Defendants	North Carolina	**30.62**
Kuziemko (2013)	Parole release and recidivism	Age at admission to prison	Georgia	**32.19**
Lofstrom, Raphael, and Grattet (2014)	Recidivism among prison released offenders	Age at release from prison	California	**36**
Stevenson (2018)	Risk assessment and pre-trial release	Pre-trial defendants	Kentucky	**33** (felony) **34** (misdemeanor)
Bureau of Justice Statistics (2013)		Felony Defendants	Large Urban Counties, 2009	**32**

Table A.2: Tests of change in covariates after introduction of 2009 changes in guidelines

	F-statistic	P-value
Covariates		
Predicted recidivism (from at-risk)	1.481849	.1919551
Predicted recidivism (from conviction)	1.733019	.1232073
Black	1.064555	.3777808
Male	1.771627	.1148739
Age at offense	1.490216	.1892112
Any previous incarceration	1.71523	.1272279
# previous cases	.2823463	.9230239
Previous incar. duration	1.867108	.0964213

Notes: This table shows the F-statistic and p-value of the Wald test of whether imbalances in punishment and covariates at each of the five discontinuities change after the introduction of the 2009 sentencing grid. The test comes from estimating Equation (4) with the location of each discontinuity defined using the *old* grid in the two years before and after the change. We then interact the 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. The F-statistics has five degrees of freedom since there are five instruments. Standard errors are clustered by individual.

Table A.3: Effect of months of incarceration on reoffending within 8 years of sentencing

	Measure of crime					
	(1) Re-incarceration	(2) Any new offense	(3) Felony	(4) Assault	(5) Property	(6) Drug
Months incar	-0.00817***	-0.00388***	-0.00241**	-0.00124	-0.00181*	-0.000283
	(0.000886)	(0.000858)	(0.000895)	(0.000771)	(0.000814)	(0.000813)
N	362989	362989	362989	362989	362989	362989
Dep. var. mean among non-incarcerated	0.552	0.593	0.457	0.142	0.253	0.290
One year effect in percentages	-17.8	-7.86	-6.34	-10.4	-8.58	-1.17
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	128.3	128.3	128.3	128.3	128.3	128.3
J stat	6.646	1.315	0.655	13.57	1.577	3.663
J stat p	0.156	0.859	0.957	0.00879	0.813	0.454
Hausman p	0.0542	0.989	0.766	0.950	0.847	0.0481
Lochner-Moretti p	0.101	0.116	0.170	0.659	0.0999	0.171

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data between 0 and 8 years of the individual's sentencing date. Standard errors are clustered by individual. Each column represents a different type of new offense (e.g., drug, property). The estimates in each column correspond to Equation (4).

Table A.4: Effect of months of incarceration on different types of probation violations within 3 years of conviction

	(1) Prob. revoke (DPS)	(2) Revoke viol.	(3) Revoke technical viol.	(4) New crime viol.	(5) Abscond viol.	(6) Drug viol.	(7) Technical viol.
Months incar	-0.0162***	-0.00991***	-0.00953***	-0.00545***	-0.00788***	-0.0101***	-0.0154***
	(0.000851)	(0.000942)	(0.000897)	(0.000869)	(0.000765)	(0.000896)	(0.00106)
N	238463	238463	238463	238463	238463	238463	238463
Dep. var. mean	0.306	0.292	0.269	0.212	0.176	0.292	0.531

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable in each column is an indicator for whether different types of probation violation occurred within 3 years of the individual's sentencing date. The coefficients reported come from 2SLS estimates of incarceration length in months instrumented with the five punishment type discontinuities, as in Equation (4). Columns 1 to 3 report estimates when the outcome is a probation revocation and Columns 4 to 7 report estimates when the outcome is a probation violation. A violation does not need to result in a revocation. The detailed information on violations that is used to calculate this estimates is available only from 2006 on, which reduces the total number of observations. Standard errors are clustered by individual.

Table A.5: Effect of months of incarceration on various reoffending outcomes within 1 year of sentencing

	Measure of crime					
	(1) Re-incarceration	(2) Any new offense	(3) Felony	(4) Assault	(5) Property	(6) Drug
Months incar	-0.0103***	-0.00878***	-0.00608***	-0.00171***	-0.00315***	-0.00212***
	(0.000474)	(0.000516)	(0.000434)	(0.000209)	(0.000273)	(0.000245)
N	516782	516782	516782	516782	516782	516782
Dep. var. mean among non-incarcerated	0.255	0.235	0.159	0.0281	0.0840	0.0768
One year effect in percentages	-48.5	-44.9	-45.8	-72.9	-45.0	-33.1
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	225.3	225.3	225.3	225.3	225.3	225.3
J stat	116.2	52.67	38.30	2.908	39.90	20.98
J stat p	3.38e-24	1.00e-10	9.71e-08	0.573	4.54e-08	0.000319
Hausman p	1.07e-17	2.28e-17	4.44e-13	2.13e-08	0.0000186	0.0144
Lochner-Moretti stat	-0.00212	-0.00264	-0.00199	-0.000819	-0.000739	-0.000271
Lochner-Moretti p	7.76e-08	1.21e-08	0.000000691	0.0000516	0.00423	0.258

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data within 1 year of the individual's sentencing date. Standard errors are clustered by individual. Each column represents a different type of new offense (e.g., drug, property). The estimates in each column correspond to Equation (4).

Table A.6: Estimates by offender and reoffending category

	Measure of crime					
	(1)	(2)	(3)	(4)	(5)	(6)
	Re-incarceration	Any new offense	Felony	Assault	Property	Drug
All offenders	-0.0138***	-0.00794***	-0.00601***	-0.00216***	-0.00327***	-0.00281***
	(0.000724)	(0.000751)	(0.000702)	(0.000435)	(0.000541)	(0.000493)
Assault offenders	-0.0131***	-0.00727***	-0.00441***	-0.00281***	-0.00241**	-0.00208**
	(0.00115)	(0.00121)	(0.00109)	(0.000793)	(0.000770)	(0.000789)
Drug offenders	-0.0152***	-0.00910***	-0.00819***	-0.00131	-0.00689***	-0.00252*
	(0.00162)	(0.00164)	(0.00160)	(0.000912)	(0.00140)	(0.00111)
Property offenders	-0.0182***	-0.0106***	-0.00967***	-0.00170*	-0.00315**	-0.00585***
	(0.00174)	(0.00171)	(0.00163)	(0.000855)	(0.00119)	(0.00137)
Other offenders	-0.0125***	-0.00593***	-0.00444***	-0.00145*	-0.00227*	-0.00161*
	(0.00118)	(0.00123)	(0.00116)	(0.000698)	(0.000890)	(0.000669)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data for each type of offense between 0 and 3 years of the individual's sentencing date. Standard errors are clustered by individual. Offender categorization refers to the focal offense for which the individual is being sentenced.

Table A.7: 2SLS estimates of incarceration length on any new offense or probation revocation within 3 years of at-risk

	2SLS			CF
	(1)	(2)	(3)	(4)
Years incar	-0.0658***	-0.0618***	-0.0128	-0.0832***
	(0.0113)	(0.0114)	(0.175)	(0.00684)
Any incarceration			-0.204	-0.0635***
			(0.133)	(0.00710)
Years incar square			0.00916	0.0163***
			(0.0438)	(0.00163)
$\hat{\nu}$				0.0634***
				(0.00672)
Any incarceration × $\hat{\nu}$				-0.0191**
				(0.00640)
Years incar × $\hat{\nu}$				-0.0121*
				(0.00511)
Years incar square × $\hat{\nu}$				-0.00118
				(0.00104)
Marginal effects of years of incarceration (in %)				
1 year incarceration effect (%)	-11.56	-10.86	-36.48	-22.91
SE	1.983	1.995	6.104	1.201
3 year incarceration effect (%)	-34.67	-32.58	-28.08	-29.29
SE	5.949	5.986	6.142	2.020
2 to 3 years incarceration effect (%)	-11.56	-10.86	5.806	-0.334
SE	1.983	1.995	8.876	0.710
Obs.	477616	477616	477616	477616
Dep. mean of non-incarcerated	0.569	0.569	0.569	0.569
Age at release FEs	No	Yes	Yes	Yes
Year of release FEs	No	Yes	Yes	Yes
J-stat (punishment type discontinuities)	25.24	25.14	4.880	2.986
J-stat p (punishment type discontinuities)	0.0000449	0.0000471	0.0872	0.0840

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows 2SLS estimates of the effects of incarceration on committing any new offense or probation revocation within 3 years of at-risk and compares them to preferred control function estimates. See the notes to Table 5 for additional details.

Table A.8: Control Function Estimates: Any new offense within 3 years of at-risk (excluding probation revocations)

	OLS	2SLS	CF			
	(1)	(2)	(3)	(4)	(5)	(6)
Any incarceration	0.0147***	-0.0651	0.000720	-0.0361***	-0.0360***	-0.0364***
	(0.00382)	(0.136)	(0.00510)	(0.00778)	(0.00779)	(0.00780)
Years incar	-0.00234	-0.0654	-0.00899*	0.00410	0.00424	0.00441
	(0.00428)	(0.178)	(0.00458)	(0.00733)	(0.00736)	(0.00737)
Years incar square	-0.00116	0.0231	-0.00111	-0.000103	-0.000189	-0.000447
	(0.000875)	(0.0442)	(0.000875)	(0.00170)	(0.00175)	(0.00176)
$\hat{\nu}$ (selection on unobserved criminality)			0.0134***	0.0394***	0.0387***	0.0387***
			(0.00329)	(0.00740)	(0.00803)	(0.00803)
Any incarceration × $\hat{\nu}$				0.00783	0.00876	0.00906
				(0.00707)	(0.00831)	(0.00832)
Years incar × $\hat{\nu}$				-0.0287***	-0.0286***	-0.0287***
				(0.00540)	(0.00544)	(0.00545)
Years incar square × $\hat{\nu}$				0.00305**	0.00307**	0.00319**
				(0.00108)	(0.00109)	(0.00109)
$\hat{\nu}^2$					-0.000356	-0.000471
					(0.00171)	(0.00171)
Selection correction for censoring						-0.000492
						(0.000715)
Marginal effects of years of incarceration (in %)						
1 year incarceration effect (%)	2.477	-13.75	-2.081	-7.127	-7.095	-7.204
SE	0.429	6.919	1.183	1.631	1.637	1.642
3 year incarceration effect (%)	-0.615	-13.67	-8.037	-5.488	-5.551	-6.042
SE	0.743	8.607	1.932	2.716	2.737	2.756
2 to 3 years incarceration effect (%)	-1.803	11.81	-3.224	0.796	0.731	0.482
SE	0.314	6.369	0.459	0.917	0.975	0.983
J-stat (punishment type discontinuities)		2.476	8.400	1.914		
J-stat p (punishment type discontinuities)		0.290	0.0780	0.167		
J-stat (all discontinuities)		36.82	46.65	34.73	32.79	29.06
J-stat p (all discontinuities)		0.00356	0.000401	0.00432	0.00502	0.0103
Obs.	397000	397000	397000	397000	397000	396907
Dep. mean of non-incarcerated	0.451	0.451	0.451	0.451	0.451	0.451
Age at release FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year of release FEs	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows estimates of several specifications of the control function approach and a comparison of the estimates to 2SLS and OLS estimates. The dependent variable is an indicator for any charges recorded in the AOC or DPS data between 0 and 3 years of the individual's release date. Standard errors are clustered by individual. The estimated control function $\widehat{\lambda}\left(X_i, Z_i^l, d\right) = \mathbb{E}\left[\nu_i | X_i, Z_i^l, D_i = d\right]$ is denoted by $\hat{\nu}$ in the table and $\hat{\nu}^2 = \mathbb{E}\left[\nu^2 | X_i, Z_i^l, D_i = d\right]$. The marginal effects show the impacts of exposure to incarceration normalized by the rate of reoffending among non-incarcerated individuals. The J-tests at the bottom of the table show model fit diagnostics. The J-test for the control function models comes from a 2SLS estimation of each specification when the endogenous variables are the $\hat{\nu}$ terms and the instruments are as specified in the table. This provides an over-identification test for the coefficients on the control function terms.

Table A.9: Heterogeneity in Control Function Estimates: Any new offense within 3 years of at-risk (excluding probation revocations)

	Population	Compliers				
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Class E	Class F	Class G	Class H	Class I
1 year incarceration effect (%)	-7.127***	-8.438***	-9.010***	-5.282***	-6.099***	-11.10***
	(1.631)	(1.960)	(2.078)	(1.314)	(1.411)	(2.635)
3 year incarceration effect (%)	-5.488*	-4.500	-5.641	-0.200	-3.563	-16.64***
	(2.716)	(3.500)	(3.615)	(2.617)	(2.484)	(3.441)
2 to 3 years incarceration effect (%)	0.796	1.775	1.547	2.201*	1.154	-2.125***
	(0.917)	(1.301)	(1.299)	(1.099)	(0.907)	(0.613)
$\sum_{d=1}^{\bar{D}} E[\nu_i \mid D_i(1) \geq d > D_i(0)] \cdot \Pr(D_i(1) \geq d > D_i(0))$ (average ν)	0	-0.184	-0.117	-0.470	-0.148	0.970
$\sum_{d=1}^{\bar{D}} d \cdot \Pr(D_i(1) \geq d > D_i(0))$ (in months of incarceration)	.	21.57	15.23	15.30	19.71	7.586
Dep. mean of non-incarcerated	0.451	0.342	0.333	0.450	0.484	0.445
Age at release FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year of release FEs	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows estimates of the marginal effects of incarceration by felony class. The dependent variable is an indicator for any charges recorded in the AOC or DPS data between 0 and 3 years of the individual's release date. All estimates are based on the same specification as in Column 4 of Table 5. All the marginal treatment effects are expressed in % terms relative to the mean reoffending rate among non-incarcerated offenders. Column 1 reports population treatment effects, i.e., when $\hat{\nu} = 0$. Columns 2-5 report estimates using the average unobserved heterogeneity ($\hat{\nu}$) of compliers in each offense class, i.e., $\sum_{d=1}^{\bar{D}} E[\nu_i|D_i(1) \geq d > D_i(0)] \cdot \omega_d$. This term is the average $\hat{\nu}$ of compliers in a given offense class. Notice that the weights ω_d are always positive and sum to one since $\omega_d \equiv \frac{\Pr(D_i(1) \geq d > D_i(0))}{\sum_{j=1}^{\bar{D}} \Pr(D_i(1) \geq j > D_i(0))}$; see the description in Section 2 for more details. Similarly $\sum_{d=1}^{\bar{D}} d \cdot \omega_d$ is the average change in exposure to incarceration due to a punishment type discontinuity expressed in terms of months of incarceration. Standard errors are clustered by individual.

Table A.10: Estimates of lower and upper bounds of the costs/value of crime

Offense category	Lower bound $			Upper bound $		
	Raw estimate	Including discounting	Reference	Raw estimate	Including discounting	Reference
Homicide	7,000,000	7,350,000	Chalfin and McCrary (2017)	9,700,000	19,205,337	Cohen et al. (2004)
Rape	142,020	149,121	Chalfin and McCrary (2017)	237,000	469,243.8	Cohen et al. (2004)
Assault	38,924	40,870.2	Chalfin and McCrary (2017)	70,000	138,595.2	Cohen et al. (2004)
Robbery	12,624	13,255.2	Chalfin and McCrary (2017)	232,000	459,344.1	Cohen et al. (2004)
Arson	38,000	128,681	Miller et al. (1996)	38,000	128,681	Miller et al. (1996)
Burglary	2,104	2,209.2	Chalfin and McCrary (2017)	25,000	49498.29	Cohen et al. (2004)
Larceny	473	497	Chalfin and McCrary (2017)	370	1,253	Miller et al. (1996)
Theft	473	497	Chalfin and McCrary (2017)	370	1,253	Miller et al. (1996)
Drug	500	990		2,544	2,945	Mueller-Smith (2015)
DWI	500	990		25,842	29,915	Mueller-Smith (2015)
Other	500	990	Cohen et al. (2004)	500	990	Cohen et al. (2004)

Notes: "Discounting" means updating the cost estimate to 2018 $, using a rate of 5% as in Mueller-Smith (2015). Offenses without a relevant cost estimate are assigned a value of $990 (in 2018 $) as was suggested by Cohen et al. (2004). The lower bounds for drug and DWI offenses were assigned in this way.

B Failure Functions and IV Estimators

B.1 Estimand using binary endogenous variable

Angrist and Imbens (1995) showed that:

$$\gamma_{\text{conviction}}(t) \equiv \frac{\mathbb{E}\left[Y_{i,t}|Z_i = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i = 0\right]}{\mathbb{E}\left[1(D_i > 0)|Z_i = 1\right] - \mathbb{E}\left[1(D_i > 0)|Z_i = 0\right]} = \beta_{\text{conviction}}(t) \cdot (1 + \kappa) \tag{B.1}$$

where

$$\kappa \equiv \frac{\sum_{l=2}^{\bar{D}} \Pr(D_i(1) \geq l > D_i(0))}{\Pr(D_i(1) \geq 1 > D_i(0))} \tag{B.2}$$

Therefore using $\mathbb{1}\{D_i > 0\}$ as the endogenous treatment yields a biased estimate of the average causal response $\beta_{\text{conviction}}(t)$; however, $\gamma_{\text{conviction}}(t)$ still has a causal interpretation as capturing a different treatment effect than the average causal response. Specifically, Equation (B.3) shows that $\gamma_{\text{conviction}}(t)$ can also be interpreted as identifying a linear combination of the *extensive* and *intensive* margin impacts of incarceration on an outcome of interest. Extensive effects are those on individuals who counterfactually would have received no incarceration sentence ($D_i(1) > D_i(0) = 0$). Similarly, intensive margin effects are the impacts of lengthening the period of incarceration for individuals who otherwise would have spent less (but not zero) time behind bars ($D_i(1) > D_i(0) > 0$).

$$\begin{aligned}\gamma_{\text{conviction}}(t) = {} & \underbrace{\mathbb{E}\left[Y_{i,t}(D_i(1)) - Y_{i,t}(0)|D_i(1) > D_i(0) = 0\right]}_{\text{Extensive margin}} + \\ & \left(\underbrace{\mathbb{E}\left[Y_{i,t}(D_i(1)) - Y_{i,t}(D_i(0))|D_i(1) > D_i(0) > 0\right]}_{\text{Intensive margin}}\right) \frac{\Pr(D_i(1) > D_i(0) > 0)}{\Pr(D_i(1) > D_i(0) = 0)}\end{aligned} \tag{B.3}$$

Notice that the weights on these two effects do not sum to one, making the estimand a linear combination of causal effects and not a weighted average. This can produce an estimand that is potentially larger than one even when the outcome is binary. However, if the instrument has no intensive margin effects (i.e., $\Pr(D_i(1) > D_i(0) > 0) = 0$), then $\gamma_{\text{conviction}}(t)$ is an estimand with a well-defined causal interpretation: it identifies the average effect of any incarceration sentence for individuals shifted to incarceration due to the instrument, i.e., extensive margin compliers. Note that the null $H_0 : \Pr(D_i(1) > D_i(0) > 0) = 0$ can be empirically examined by testing the following necessary condition that most hold if the null is true: $H_0 :\ \Pr(D_i(1) \geq 1 > D_i(0)) \geq \Pr(D_i(1) \geq d > D_i(0))\ \forall d > 1$. This is a necessary condition, and not a sufficient condition, for the null to be satisfied.

B.2 Failure function as cumulative hazards

The failure function $Y_{i,t}(d)$ of reoffending within t periods from the date of conviction can be written recursively, which will be convenient for what follows:

$$Y_{i,t} = Y_{i,t-1} + y_{i,t} \cdot (1 - Y_{i,t-1}) \tag{B.4}$$

where y_{it} is an indicator for whether individual i reoffends *at* period t from the date of sentencing. The recursive formulation implies that analogous Wald estimates of the failure function for t periods from conviction result in:

$$\begin{aligned}
&\beta_{\text{conviction}}(t) = \frac{\mathbb{E}\left[Y_{i,t}|Z_i = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i = 0\right]}{\mathbb{E}\left[D_i|Z_i = 1\right] - \mathbb{E}\left[D_i|Z_i = 0\right]} \\
&= \beta_{\text{conviction}}(t-1) + \frac{\mathbb{E}\left[y_{i,t} \cdot (1 - Y_{i,t-1})\,|Z_i = 1\right] - \mathbb{E}\left[y_{i,t} \cdot (1 - Y_{i,t-1})\,|Z_i = 0\right]}{\mathbb{E}\left[D_i|Z_i = 1\right] - \mathbb{E}\left[D_i|Z_i = 0\right]} \\
&= \beta_{\text{conviction}}(t-1) + \beta^{\text{hazard}}_{\text{conviction}}(t)
\end{aligned}$$

where the above derivations also hold for Wald estimates of the effects of incarceration on reoffending from the date of release.

The second term in Equation (B.5) is related to treatment effects on a discrete-time hazard. That is, it measures the effect of a incarceration on the probability of offending at period t after conviction, conditional on having not offended previously, i.e., having survived up until that point. We can thus express the Wald estimator for $\beta_{\text{conviction}}(t)$ as the sum of individual $\beta^{\text{hazard}}_{\text{conviction}}(t)$:

$$\beta_{\text{conviction}}(t) = \sum_{l=1}^{t} \beta^{\text{hazard}}_{\text{conviction}}(l) \tag{B.5}$$

In our empirical specifications, in addition to estimating the effects of incarceration on $Y_{i,t}$ for a particular t (e.g., three years), we estimate effects for $t \in (0, 60)$, where t is measured in months. The *slope* of these estimates, i.e., the difference from t to $t+1$, represents the treatment effect on hazards $\beta^{\text{hazard}}_{\text{conviction}}(t)$.

C IV using reoffending from at-risk and time-varying controls

In this appendix, we discuss the estimands that are recovered by 2SLS estimators of the effects of incarceration length (D_i) on reoffending from at-risk ($Y_{i,t+D_i}$). To identify the effects of incarceration on criminal behavior, free of any incapacitation effects, it is common to measure reoffending from the date an individuals is back in the community and is at-risk to reoffend. The estimand that is now recovered by 2SLS, without adjusting for any time-varying controls, is

$$\begin{aligned}\beta_{\text{at-risk}}(t) &\equiv \frac{\mathbb{E}\left[Y_{i,t+D_i}|Z_i=1\right]-\mathbb{E}\left[Y_{i,t+D_i}|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]} \\ &= \sum_{j=1}^{\bar{D}} \omega_j \mathbb{E}\left[Y_{i,t+d}(d)-Y_{i,t+d-1}(d-1)|D_i(1)\geq d > D_i(0)\right]\end{aligned} \tag{C.1}$$

The proof the above equality follows almost directly from Angrist and Imbens (1995). To see this we express reoffending from at-risk in terms of potential outcomes. Let $\lambda_{i,d}(Z_i)=1(D_i(Z_i)\geq d)$ denote an indicator for being incarcerated for at least d months, which is a function of whether the individual is to the right ($Z_i=1$) or the the left of the discontinuity ($Z_i=0$). When measuring reoffending from at-risk the potential outcomes are of the following form $Y_{i,t+D_i(Z_i)}(D_i(Z_i))$, i.e., the instrument assignment influences both the length of incarceration and the number of months from conviction that will be used to measure reoffending. Now observed reoffending within t months from at-risk $Y_{i,t+D_i}$ can expressed as

$$Y_{i,t+D_i} = Z_i\left[\sum_{d=1}^{\bar{D}} Y_{i,t+d}(d)\left(\lambda_{i,d}(1)-\lambda_{i,d+1}(1)\right)\right] + (1-Z_i)\left[\sum_{d=1}^{\bar{D}} Y_{i,t+d}(d)\left(\lambda_{i,d}(0)-\lambda_{i,d+1}(0)\right)\right] \tag{C.2}$$

From this point the proof is identical to the proof of Theorem 1 in Angrist and Imbens (1995).

The estimand identified by $\beta_{\text{at-risk}}(t)$ is hard to interpret since it includes adjusting the window of time in which reoffending is measured based on the endogenous treatment of interest D_i. Another concern is that the incarcerated and non-incarcerated offenders will now vary in observable and unobservable time-varying factors. For example, the age of the offender at time zero, i.e., at the point in time that we start to measure reoffending.

To illustrate the difficulties in identifying behavioral effects using $\beta_{\text{at-risk}}(t)$, we present the following example which builds directly on results from Lochner and Moretti (2015) on the properties of 2SLS estimator for treatments with multiple levels. Consider the following causal model

$$\mathbb{E}\left[Y_{i,t+D_i}\right] = X_i'\alpha + W_{i,D_i}'\gamma + \sum_{i=d}^{\bar{D}} \gamma_d \mathbb{1}\left\{D_i \geq d\right\} \tag{C.3}$$

where W_{i,D_i} are time-varying factors such as age at release. When not adjusting for time-varying controls $\beta_{\text{at-risk}}(t)$ recovers

$$\sum_{j=1}^{\bar{D}} \omega_j \gamma_d + \sum_{j=1}^{\bar{D}} \omega_j \left(W_{i,d}-W_{i,d-1}\right)'\gamma \tag{C.4}$$

However, when adjusting W_{i,D_i} in the 2SLS specification $\beta_{\text{at-risk}}(t)$ recovers

$$\sum_{j=1}^{\bar{D}} \tilde{\omega}_j \gamma_d, \quad \tilde{\omega}_j = \frac{\Pr(D_i \geq j)\mathbb{E}\left[\xi_i | D_i \geq j\right]}{\sum_{l=1}^{\bar{D}} Pr(D_i \geq l)\mathbb{E}\left[\xi_i | D_i \geq l\right]} \tag{C.5}$$

where ξ_i is the residual from projecting Z_i on W_{i,D_i}, i.e., from the projection $Z_i = W'_{i,D_i}\alpha + \xi_i$. The estimand in Equation (C.5) is a linear combination of causal effects; however, the $\tilde{\omega}_j$ weights can have negative values—ruling out the option of interpreting the estimand as a weighted average of causal effects. Since some of the weights can potentially be grater than one and also some can be positive while others negative, it is not clear what is the interpretation of the object that is identified when including adjustments for time-varying confounders in the 2SLS model.

Moreover, the treatment effects can also be heterogeneous, in addition to being non-linear. For example, consider the scenario that the $\gamma_{i,d}$ are random coefficients that potentially vary by individual. Now the 2SLS estimator will recover:

$$\sum_{j=1}^{\bar{D}} \frac{\Pr(D_i \geq j)\mathbb{E}\left[\xi_i \gamma_{i,d} | D_i \geq j\right]}{\sum_{l=1}^{\bar{D}} Pr(D_i \geq l)\mathbb{E}\left[\xi_i \gamma_{i,d} | D_i \geq l\right]} \tag{C.6}$$

This estimand is even harder to interpret in causal terms, since it involves the correlation between treatment effects and the residuals from the projection of Z_i of $W_{i,d}$. Now the 2SLS estimand can no longer be represented as a linear combination of causal effects.

The above example illustrates that non-parametrically identifying behavioral responses is difficult when using only an IV estimator. To formally layout identification results for behavioral effects separately from any time-varying confounders, we present a control function approach (Section 7) that makes additional parametric restriction on the data generating process, mainly additive separability in the outcome equation, but provides a semi-structural framework to identify the behavioral effects of incarceration.

D Sentencing grids in North Carolina

***** Effective for Offenses Committed on or after 12/1/95 *****

FELONY PUNISHMENT CHART

PRIOR RECORD LEVEL

OFFENSE CLASS	I 0 Pts	II 1-4 Pts	III 5-8 Pts	IV 9-14 Pts	V 15-18 Pts	VI 19+ Pts	
A	**Death or Life Without Parole**						
B1	**A**	**A**	**A**	**A**	**A**	**A**	**DISPOSITION**
	240 - 300	*288 - 360*	*336 - 420*	*384 - 480*	*Life Without Parole*	*Life Without Parole*	*Aggravated Range*
	192 - 240	**230 - 288**	**269 - 336**	**307 - 384**	**346 - 433**	**384 - 480**	**PRESUMPTIVE RANGE**
	144 - 192	*173 - 230*	*202 - 269*	*230 - 307*	*260 - 346*	*288 - 384*	*Mitigated Range*
B2	**A**	**A**	**A**	**A**	**A**	**A**	
	157 - 196	*189 - 237*	*220 - 276*	*251 - 313*	*282 - 353*	*313 - 392*	
	125 - 157	**151 - 189**	**176 - 220**	**201 - 251**	**225 - 282**	**251 - 313**	
	94 - 125	*114 - 151*	*132 - 176*	*151 - 201*	*169 - 225*	*188 - 251*	
C	**A**	**A**	**A**	**A**	**A**	**A**	
	73 - 92	*100 - 125*	*116 - 145*	*133 - 167*	*151 - 188*	*168 - 210*	
	58 - 73	**80 - 100**	**93 - 116**	**107 - 133**	**121 - 151**	**135 - 168**	
	44 - 58	*60 - 80*	*70 - 93*	*80 - 107*	*90 - 121*	*101 - 135*	
D	**A**	**A**	**A**	**A**	**A**	**A**	
	64 - 80	*77 - 95*	*103 - 129*	*117 - 146*	*133 - 167*	*146 - 183*	
	51 - 64	**61 - 77**	**82 - 103**	**94 - 117**	**107 - 133**	**117 - 146**	
	38 - 51	*46 - 61*	*61 - 82*	*71 - 94*	*80 - 107*	*88 - 117*	
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*	

A – Active Punishment **I – Intermediate Punishment** **C – Community Punishment**
Numbers shown are in months and represent the range of minimum sentences

Revised: 08-04-95

***** Effective for Offenses Committed on or after 12/1/09 *****

FELONY PUNISHMENT CHART

PRIOR RECORD LEVEL

OFFENSE CLASS

	I 0-1 Pt	II 2-5 Pts	III 6-9 Pts	IV 10-13 Pts	V 14-17 Pts	VI 18+ Pts
A	Death or Life Without Parole					
B1	**A**	**A**	**A**	**A**	**A**	**A**
	240 - 300	*276 - 345*	*317 - 397*	*365 - 456*	*Life Without Parole*	*Life Without Parole*
	192 - 240	**221 - 276**	**254 - 317**	**292 - 365**	**336 - 420**	**386 - 483**
	144 - 192	*166 - 221*	*190 - 254*	*219 - 292*	*252 - 336*	*290 - 386*
B2	**A**	**A**	**A**	**A**	**A**	**A**
	157 - 196	*180 - 225*	*207 - 258*	*238 - 297*	*273 - 342*	*314 - 393*
	125 - 157	**144 - 180**	**165 - 207**	**190 - 238**	**219 - 273**	**251 - 314**
	94 - 125	*108 - 144*	*124 - 165*	*143 - 190*	*164 - 219*	*189 - 251*
C	**A**	**A**	**A**	**A**	**A**	**A**
	73 - 92	*83 - 104*	*96 - 120*	*110 - 138*	*127 - 159*	*146 - 182*
	58 - 73	**67 - 83**	**77 - 96**	**88 - 110**	**101 - 127**	**117 - 146**
	44 - 58	*50 - 67*	*58 - 77*	*66 - 88*	*76 - 101*	*87 - 117*
D	**A**	**A**	**A**	**A**	**A**	**A**
	64 - 80	*73 - 92*	*84 - 105*	*97 - 121*	*111 - 139*	*128 - 160*
	51 - 64	**59 - 73**	**67 - 84**	**78 - 97**	**89 - 111**	**103 - 128**
	38 - 51	*44 - 59*	*51 - 67*	*58 - 78*	*67 - 89*	*77 - 103*
E	**I/A**	**I/A**	**A**	**A**	**A**	**A**
	25 - 31	*29 - 36*	*33 - 41*	*38 - 48*	*44 - 55*	*50 - 63*
	20 - 25	**23 - 29**	**26 - 33**	**30 - 38**	**35 - 44**	**40 - 50**
	15 - 20	*17 - 23*	*20 - 26*	*23 - 30*	*26 - 35*	*30 - 40*
F	**I/A**	**I/A**	**I/A**	**A**	**A**	**A**
	16 - 20	*19 - 23*	*21 - 27*	*25 - 31*	*28 - 36*	*33 - 41*
	13 - 16	**15 - 19**	**17 - 21**	**20 - 25**	**23 - 28**	**26 - 33**
	10 - 13	*11 - 15*	*13 - 17*	*15 - 20*	*17 - 23*	*20 - 26*
G	**I/A**	**I/A**	**I/A**	**I/A**	**A**	**A**
	13 - 16	*14 - 18*	*17 - 21*	*19 - 24*	*22 - 27*	*25 - 31*
	10 - 13	**12 - 14**	**13 - 17**	**15 - 19**	**17 - 22**	**20 - 25**
	8 - 10	*9 - 12*	*10 - 13*	*11 - 15*	*13 - 17*	*15 - 20*
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

A – Active Punishment **I – Intermediate Punishment** **C – Community Punishment**
Numbers shown are in months and represent the range of <u>minimum</u> sentences

Revised: 08-31-09

E Independent competing risks and non-parametric bounds

Reincarceration due to technical probation revocations can bias incarceration effects estimates in two ways. First, if revocations mask genuine criminal activity, not counting them as reoffending may artificially deflate reoffending rates in the probation (and thus control) population. Second, even if technical revokes are not associated with actual crimes, revoked individuals may have otherwise committed crimes in the future. Since these individuals go to prison, overall offending in the control population will go down. If those revoked are also higher risk on average, the remaining control units at-risk to reoffend may be positively selected, exacerbating the problem.

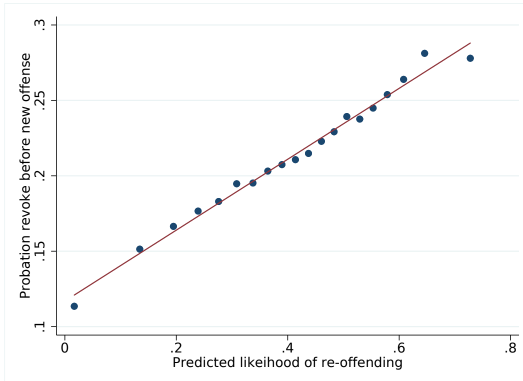
When probation revocation occur randomly, such censoring is not an issue, since reoffending rates conditional on not having probation revoked before committing a new offense provide an unbiased estimate of the untreated reoffending rate. However, individuals' likelihood of revocation can be correlated with their likelihood of reoffending, i.e., that higher risk offenders are more likely to be revoked, which implies that reoffending rates conditional on no revocation are biased towards zero. Supporting this possibility, Appendix Figure E.1 shows that there is a strong positive correlation between probation revocations and the predicted likelihood of committing a new offense within three years from being at-risk among the non-incarcerated offenders.[42]

Nevertheless, estimates assuming probation revocations and reoffending are uncorrelated may provide a plausible upper (most crime increasing) bound for the effects of incarceration. We present these estimates in Appendix Figure E.2, which adds a purple line representing estimates in a sample that drop observations with a technical probation revocation prior to committing a new offense. This line falls between the red (only new offenses) and maroon (new offense or revoke) colored lines both when measuring reoffending from conviction. The regular and independent risks estimates of committing any new offense within t months from conviction substantially differ in the first years post-conviction, but over time they converge to almost the same value. This is what we would expect if the primary impact of incarceration (including as a result of probation revocations) comes through incapacitation and the behavioral effects are crime reducing but small. The estimated incarceration effects, under an independent risks assumption, are larger in magnitude for a variety of types of new offenses (Appendix Table E.1). For example, the effect of a year of incarceration on new assault offenses changes from -2.59 p.p. (↓37.6%) to -3.42 p.p. (↓47.1%).

Finally, a completely non-parametric approach to the issue of probation revocations is to construct worst case bounds, also known as Peterson (Peterson, 1976) bounds (see Appendix F for details). The estimated bounds are tight enough to be informative and show incarceration has crime reducing effects on reoffending within three years from conviction (Appendix Table F.1).

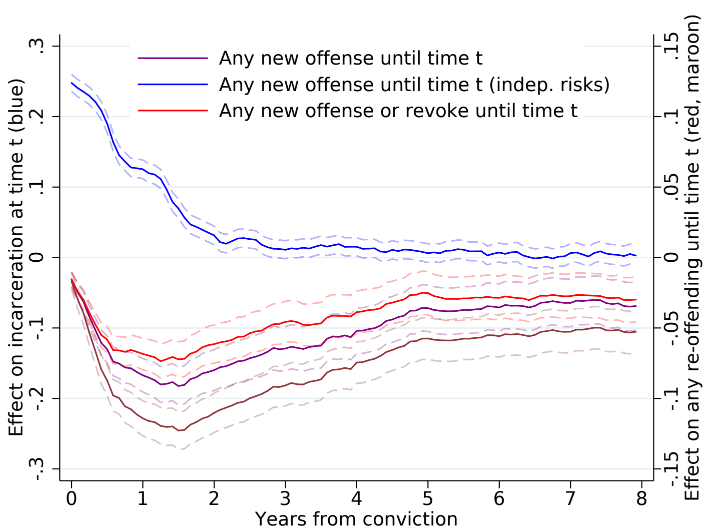
[42] It is important to note that this is a descriptive correlation only. We estimated an OLS model of committing a new offense within three years from at-risk on control variables that include criminal history and demographic information (e.g., age) using only individuals who have not been incarcerated and measured reoffending within three years of release. The figure plots only the incarcerated group, which does not include any of the observations used to construct the predictions for committing a new offense.

Figure E.1: The relationship between technical probation revocation prior to a new offense and the predicted likelihood of committing a new offense within three years



Notes: Figure displays the relationship between the predicted likelihood of committing a new offense and the likelihood of getting a technical probation revocation prior to committing a new offense. Only individuals who have not been sentenced to incarceration are shown in the graph. To avoid over-fitting issues, the model for the predicted likelihood of reoffending was estimated using only individuals who have been incarcerated and are not used in the figure. The line represents the OLS regression fit conditional on the running variables in Equation (4), i.e., trends in prior record points within a prior record level (columns) and an offense felony class (rows).

Figure E.2: Independent risks: Reduced form estimates of any incarceration *at* period t and of *any* reoffending up to period t from conviction



Notes: This figure shows reduced form estimates of being to the right of a punishment type discontinuity on several different outcomes of interest. All outcomes/measures are with respect to the conviction date. The blue line (left y-axis) represents the the reduced form effect on an indicator for spending any positive amount of time behind bars *at* month t from conviction. The red color line (right y-axis) reports the reduced form effects on committing any new offense *until* month t, and the maroon color line (right y-axis) the estimates when also including probation revocations as offending. The purple line represents estimates on committing a new offense until period t under independent risks when dropping observations in which a technical probation revocation occurred before committing a new offense. The purple line shows estimates under an independent risks assumption, when dropping observations in which a technical probation revocation occurred before committing a new offense. Standard errors are clustered by individual. See also the notes in Figure 6 for further details on the estimation.

Table E.1: Independent risks: 2SLS estimates of length of incarceration effects on different type of reoffending (by type of crime) within three years of sentencing

	Measure of crime					
	(1) Re-incarceration	(2) Any new offense	(3) Felony	(4) Assault	(5) Property	(6) Drug
Months incar	-0.00996***	-0.0105***	-0.00775***	-0.00285***	-0.00403***	-0.00358***
	(0.000773)	(0.000860)	(0.000806)	(0.000502)	(0.000620)	(0.000571)
N	411246	411246	411246	411246	411246	411246
Dep. var. mean among non-incarcerated	0.314	0.451	0.327	0.0726	0.172	0.178
One year effect in percentages	-38.1	-27.9	-28.5	-47.1	-28.1	-24.2
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	161.1	161.1	161.1	161.1	161.1	161.1
J stat	31.81	6.028	7.203	2.712	5.001	11.38
J stat p	0.00000209	0.197	0.126	0.607	0.287	0.0226
Hausman p	0.00000725	0.00000165	0.000129	0.000621	0.0476	0.115
Lochner-Moretti stat	-0.00152	-0.00280	-0.00253	-0.00134	-0.00124	-0.000597
Lochner-Moretti p	0.0426	0.000788	0.00134	0.00650	0.0424	0.291

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data between 0 and three years of the individual's sentencing date. Observation in which a probation revocation occurred prior to a new offense have been dropped according to the independent risks assumption. Standard errors (in parentheses) are clustered by individual. Each column represents a different type of new offense (e.g., drug, property). The estimates in each column are from Equations (5) and (4). The Lochner-Moretti statistics and p-values are a generalization of the standard Hausman test of endogeneity to an ordered treatment with multiple levels and potentially non-linear effects as is described in Lochner and Moretti (2015).

F Peterson (1976) Bounds for Censoring due to Technical Probation Revocations

In this appendix, we discuss non-parametric bounds that account for the effects of technical probation revocations on estimates of incarceration effects. We begin by describing the bounds and then present several different types of estimates.

The bounds can be derived by re-defining the outcome $Y_{i,t}$ under different assumptions on the unobserved correlation between reoffending and having a probation revocation to obtain bounds on incarceration estimates that are analogous to similar bounds in the competing risks literature (Peterson, 1976). Define:

$$Y_{i,t}^{ub} = \mathbb{1}\left[\sum_{l=0}^{t} y_{i,l}\left(1 - R_{i,t}\right) > 0\right] \tag{F.1}$$

Intuitively, $Y_{i,t}^{ub}(d)$ forms an upper bound by counting offenses that occur before the individual has a probation revocation, i.e., $R_{i,t} = 1$. Effectively, it assumes that if the competing revocation risk occurs before a new offense, the individual will never commit another new offense. Another interpretation is that individuals whose probation is revoked are the ones *least* likely to reoffend. If they would have not been revoked they would have not reoffended. In our setting, this bound measures reoffending such that incarceration is the most crime increasing relative to probation.

A lower bound can be derived by doing the opposite, counting either as failure. Effectively, this assumes that individuals who are revoked would have failed with a new offense in the same period. This is equivalent to counting probation revocations as reoffending. Define the lower bound as:

$$Y_{i,t}^{lb} = \mathbb{1}\left[\left(\sum_{l=0}^{t} y_{i,l} + R_{i,t}\right) > 0\right] \tag{F.2}$$

Table F.1: Peterson (1976) bounds on the effects of incarceration on committing a new offense within three years of conviction

	(1) Naive estimate	(2) Lower bound	(3) Upper bound	(4) Independent risks
Months incap	-0.00794***	-0.0133***	-0.00584***	-0.0105***
	(0.000751)	(0.000771)	(0.000749)	(0.000860)
N	491135	491135	491135	411246
Dep. var. mean	0.417	0.526	0.364	0.434
Controls	Yes	Yes	Yes	Yes
Controls criminal history	Yes	Yes	Yes	Yes

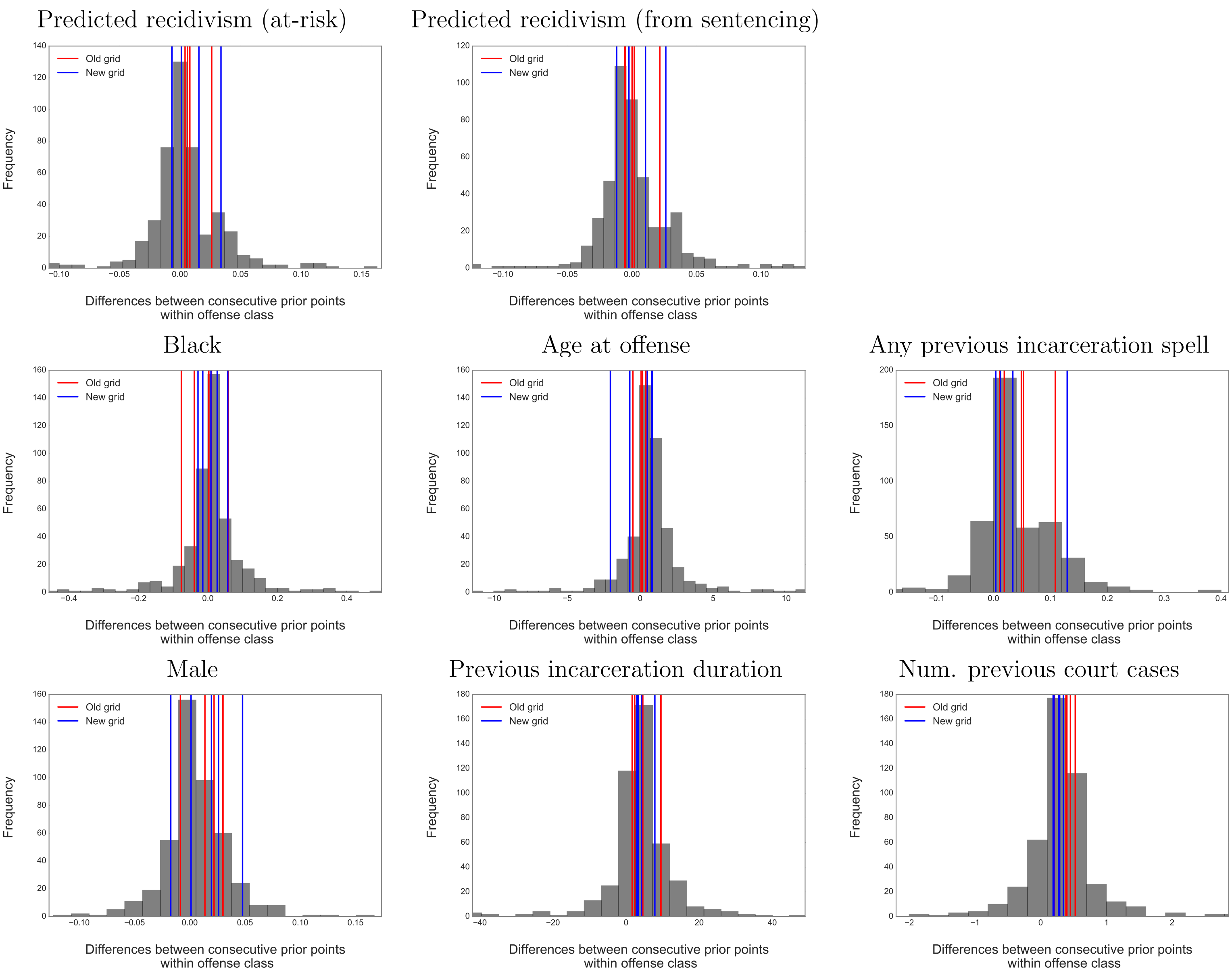
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data between 0 and three years of the individual's sentencing date. Standard errors are clustered by individual. Each column in the table represents a 2SLS coefficient of the effects of incarceration length on committing a new offense within three years of sentencing.

G Tests of instrument validity

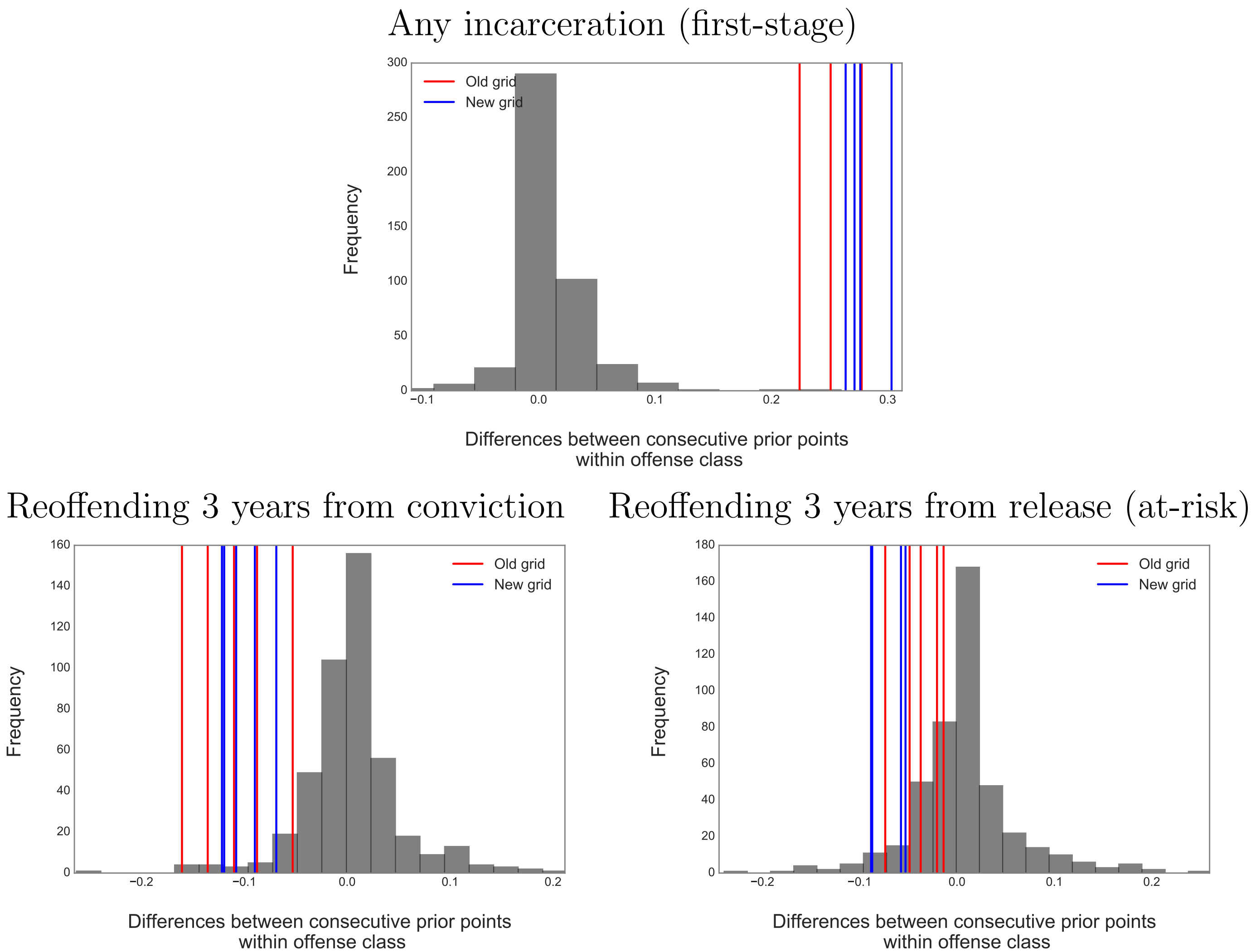
This appendix includes additional figures and tables that present evidence in support of the validity of the instrumental variables. The figures and tables are discussed in the main text of the paper.

Figure G.1: Difference in covariates before and after punishment type discontinuities relative to differences between consecutive prior points without a punishment type change



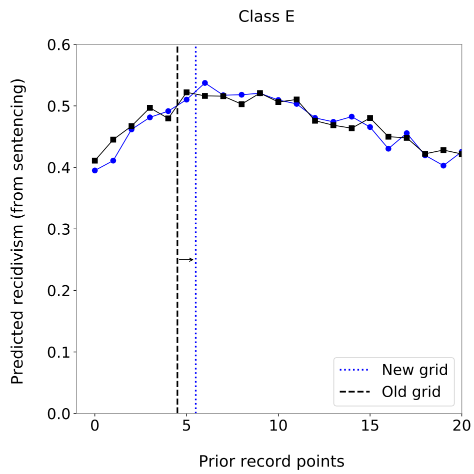
Notes: This figure tests for imblanaces in covariates (pre-conviction characteristics) at the discontinuities in punishment relative to any transition across prior points in which there is no change in punishment type. The figure plots the distribution of the difference in the mean values of a given covariate (e.g., male, black) between two consecutive prior points by felony class and before and after the grid changes in 2009. The red (or blue) lines indicate the differences at prior points transitions with a punishment type discontinuity using date before (after) the 2009 grid changes. The figure includes four different covariates, the distribution of each is plotted separately. The covariates in the figure are an indicator for whether the offender is black, the age at the time the offense took place, the predicted recidivism (i.e., reoffending) risk from at-risk and from conviciton. Since there are many important pre-treatment covariates, we make use of this predicted reoffending (risk) score that is calculated by regressing reoffending on all the pre-treatment covariates (using only non-incarcerated offenders) and fitting predicted values to all offenders. Summarizing imbalance by the covariates' relationship to the outcome surface is a common methodology in the literature Bowers and Hansen (2009) and Card et al. (2015).

Figure G.2: Difference in incarceration and reoffending before and after punishment type discontinuities relative to differences between any two consecutive prior points without a punishment type discontinuity between them

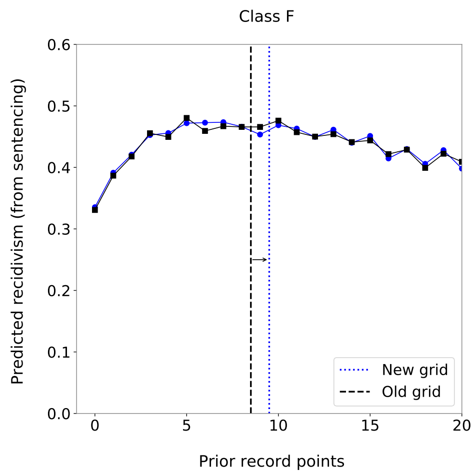


Notes: This figure illustrates the variation caused by the discontinuities in incarceration exposure (first-stage) and reoffending. The figure plots the distribution of the difference in the mean values of a given outcome (e.g., any initial incarceration, any reoffending within 3 years) between two consecutive prior points by felony class and before and after the grid changes in 2009. The red (or blue) lines indicate the differences at prior points transitions with a punishment type discontinuity using date before (after) the 2009 grid changes. The reoffending measure in the figure is any new offense or probation revocation.

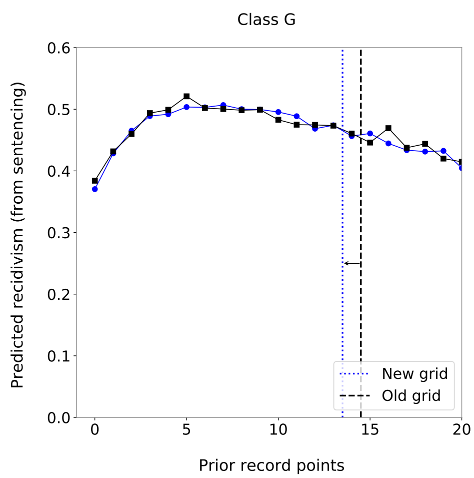
Figure G.3: Predicted recidivism score does not vary due to 2009 changes in the location of discontinuities in sentencing guidelines



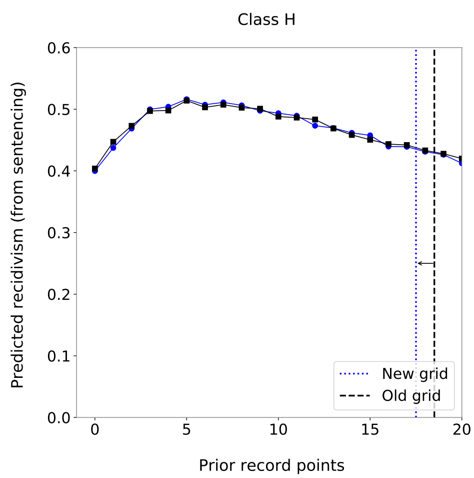
(a) Class E



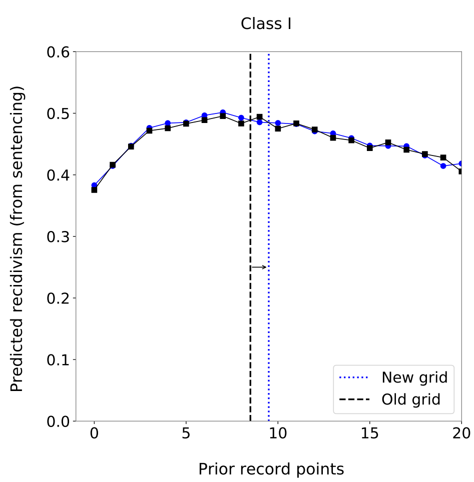
(b) Class F



(c) Class G



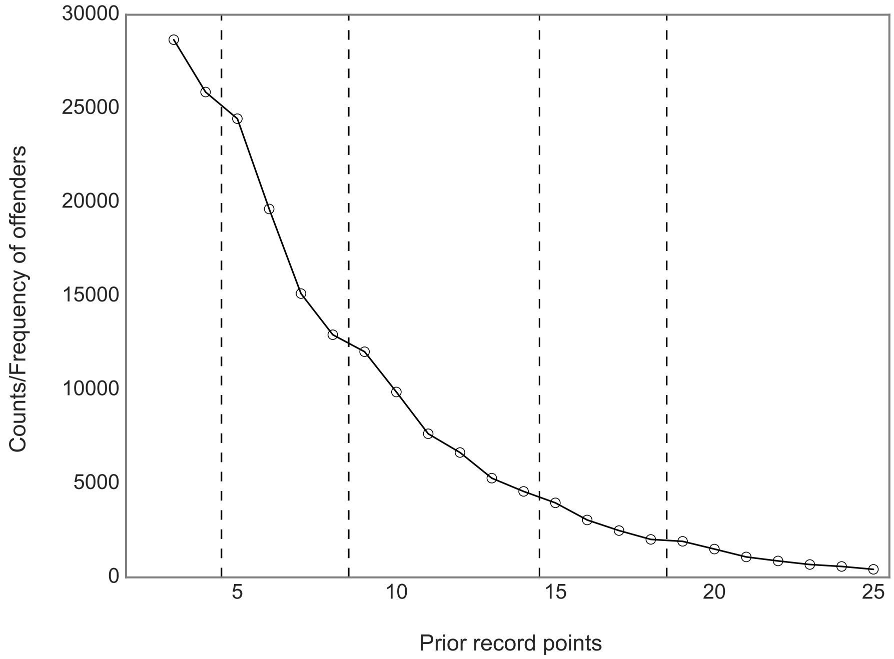
(d) Class H



(e) Class I

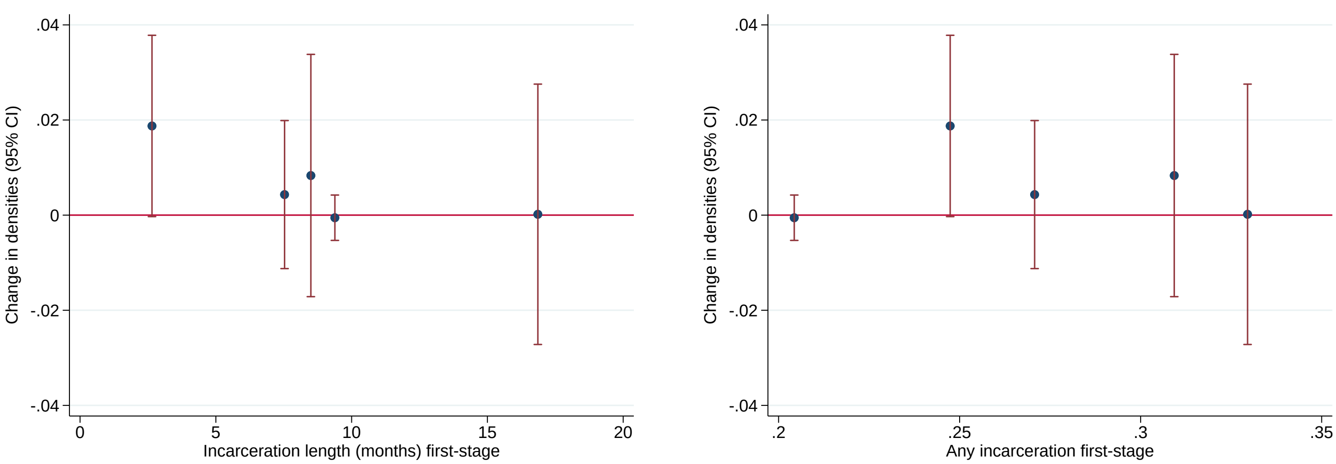
Notes: The x-axis in all plots is the number of prior record points. The y-axis reports the offender's average predicted recidivism score. The black line represents the average predicted recidivism score prior to the 2009 reform and the blue line the predicted recidivism score after the reform. The plots demonstrate how the 2009 changes in the location of discontinuities in the sentencing grid do not lead to any discontinuities in the predicted recidivism score. The old grid refers to the sentencing grid between 1996 to 2013 (see Appendix D), and the new grid refers to the sentencing from 2009 to 2011 (see Appendix A). The location of the discontinuities in the punishment type and severity did not change since the 2009 reform to the present, although changes within the grid have been made.

Figure G.4: Distribution of offenders across prior record points



Notes: The x-axis in all plots is the number of prior record points. The y-axis show the mean age of offenders at the time the offense was committed. The figure present only offenses that took place between 1995 and 2009 and have been sentenced under the sentencing grid that applied for offenses committed between 1995 to 2009, see Appendix D for the official grid. In 2009 the guidelines changes and the discontinuities shifted by one prior points either to the left or to the right, see Appendix D. The figure for offenses that took place after 2009 looks very similar and the density of individuals also varies smoothly across between prior record levels.

Figure G.5: VIV of punishment severity first-stage and reduced-form coefficients of changes in density (count of observations) at the different discontinuities



H Heterogeneity by discontinuity

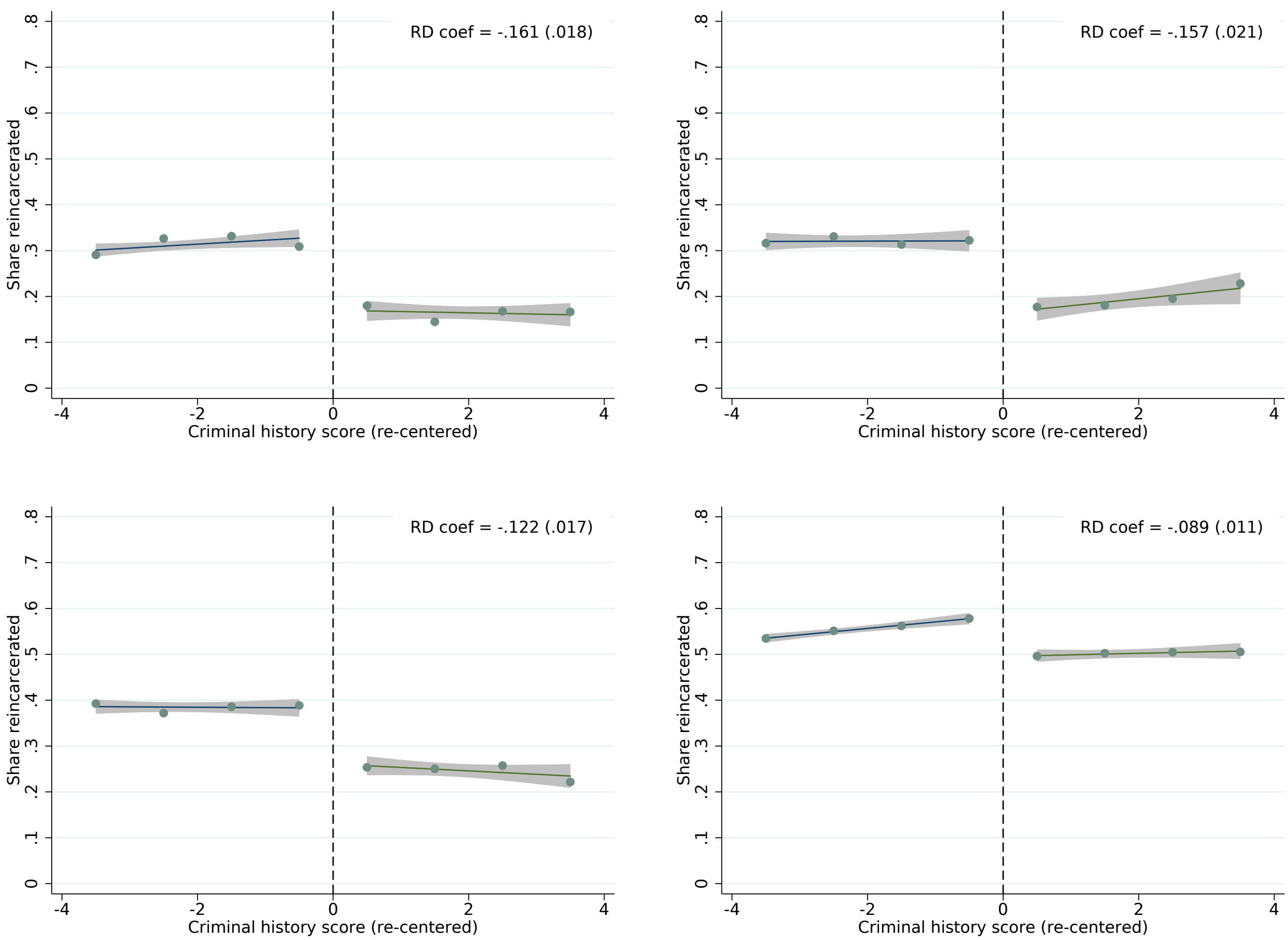
In this appendix, we explore heterogeneity by felony class and report estimates of reduced form figures that are analogous to Figure 6 for each felony class separately. As noted in the main text, the reduced form results combine and average the effects of crossing multiple discontinuities. Because each discontinuity applies to different offenders, has a different first stage, and has different mean compliance rates to the left and the right of the threshold, each may also capture treatment effects for different complier populations. Because each instrument also shifts exposure to different amounts of incarceration, the reduced forms may also vary because they capture different weighted averages of the same incremental treatment effects (see Equation (1)).

In Appendix I, we present a first evidence of such heterogeneity by estimating the characteristics of compliers. Since there are varying levels of treatment, there are also multiple classes of compliers, i.e., $\mathbb{E}\left[X_i|D_i(1) \geq d > D_i(0)\right]$ for each level of d. In the appendix, we extend Abadie (2002) to the case of a treatment with multiple levels. The results show that compliers do differ substantially across each instrument with respect to both their demographic characteristics and their criminal histories.

Appendix Figures H.6 and H.7 show the main reduced form estimates by felony class. Panel (a) plots documents effects on incarceration and reoffending at the monthly level. The patterns in all the classes look similar, although there is substantial variation in duration of incarceration. For example, in class I, the instruments stop being predictive of incarceration status one year from conviction; however, in class E it takes over four years. Nevertheless, in all classes there is a reduction in the period-by-period offending rates while the instruments are predictive of incarceration status and afterwards no visible differences in monthly reoffending rates.

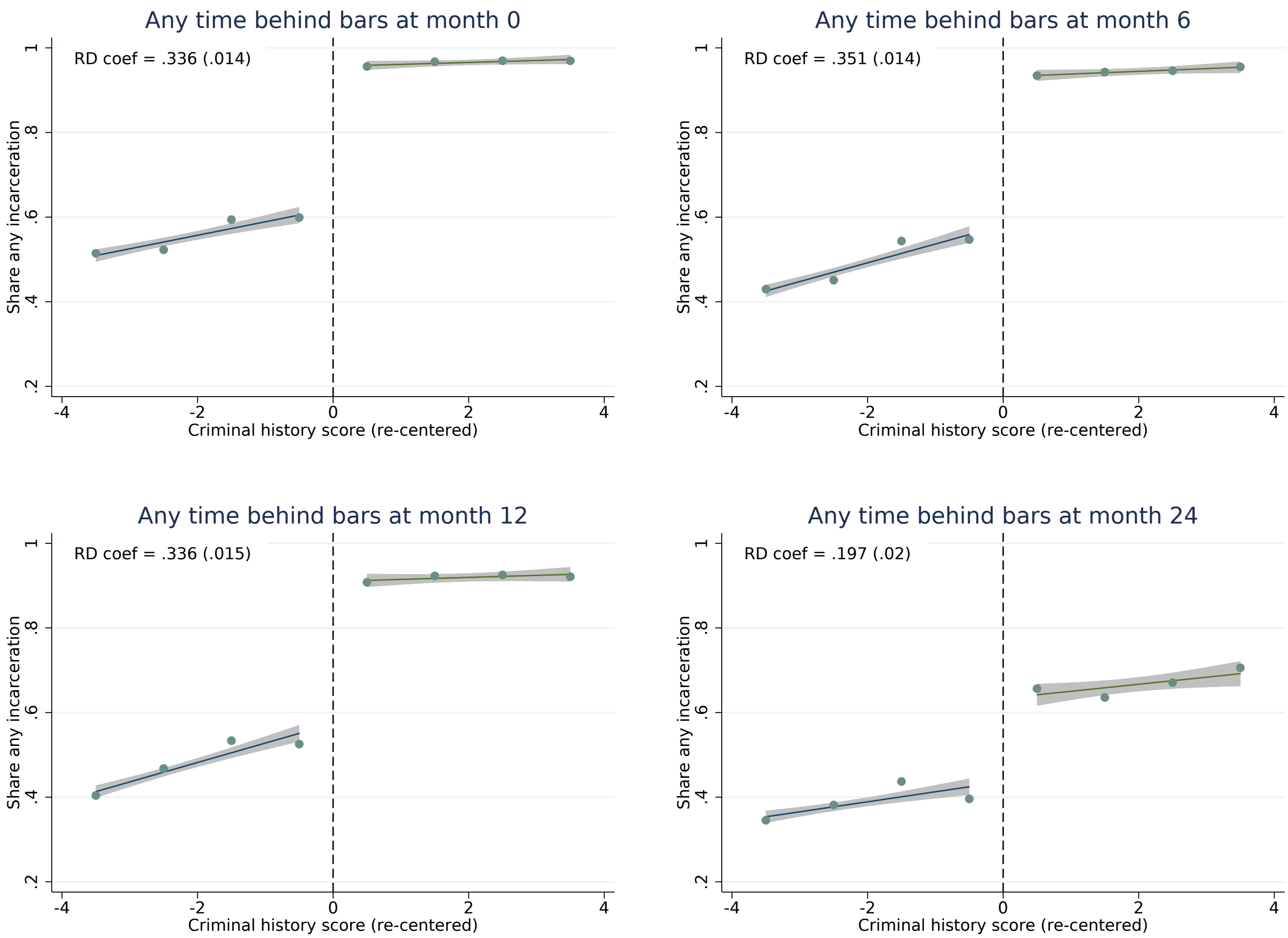
Panel (b) plots shows that although there is substantial heterogeneity in the magnitude of the incapacitation effects, the impacts on any reoffending in the long term show either a zero effect (e.g., class I) or permanent reduction in some classes (e.g., E or F). It is interesting to note that the reduced forms with the largest permanent reductions in offending also have the longest incarceration treatments. Thus while no class shows incarceration ever increases offending post-release, there is some suggestive evidence that only longer sentences persistently reduce it.

Figure H.1: Share reincarcerated within three years of conviction (offenses from felony classes E,G,H, and I)



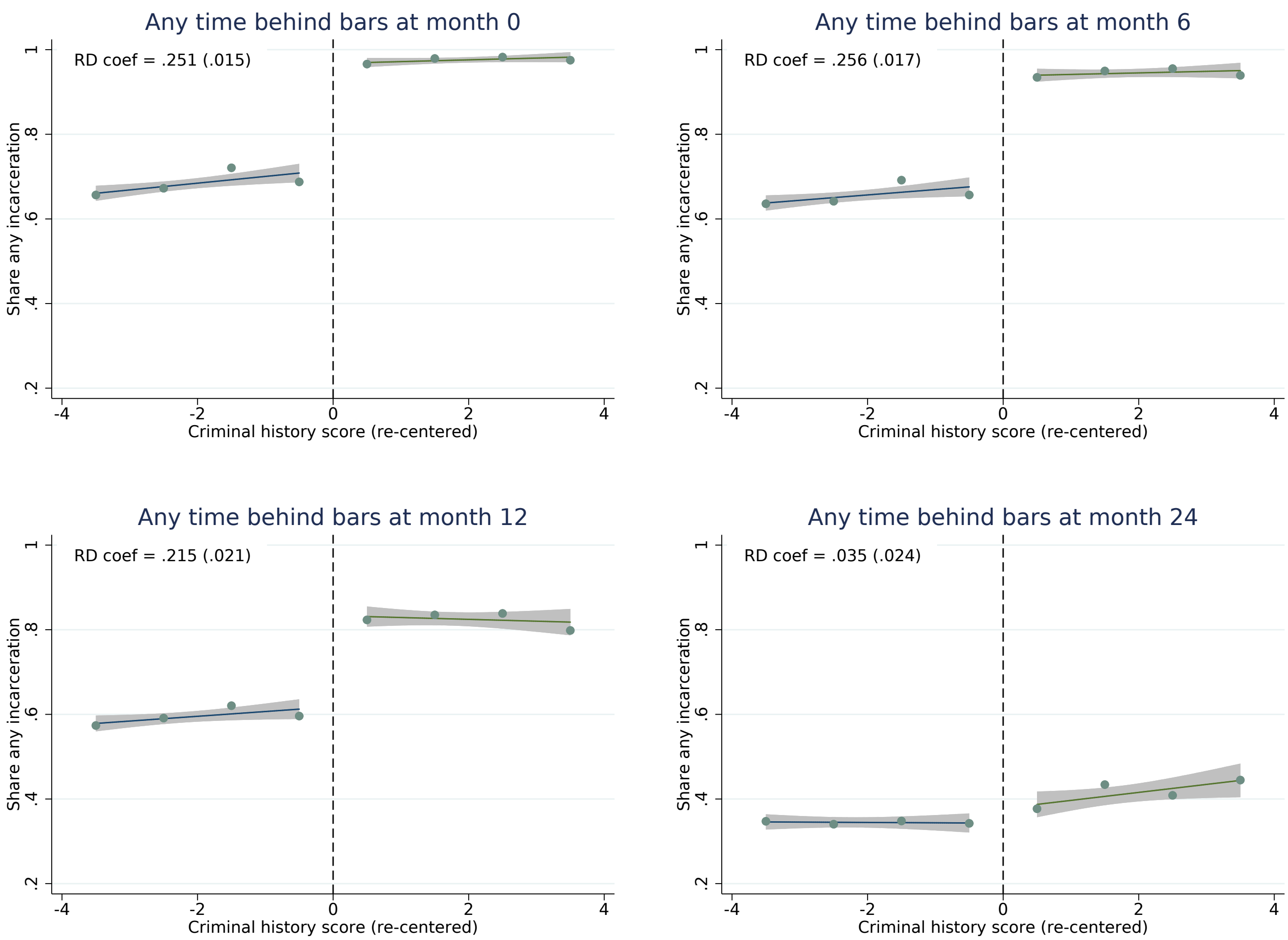
Notes: See notes to Figure 4.

Figure H.2: Dynamic differences in incarceration status *at* a given month after conviction (class E felony offenses)



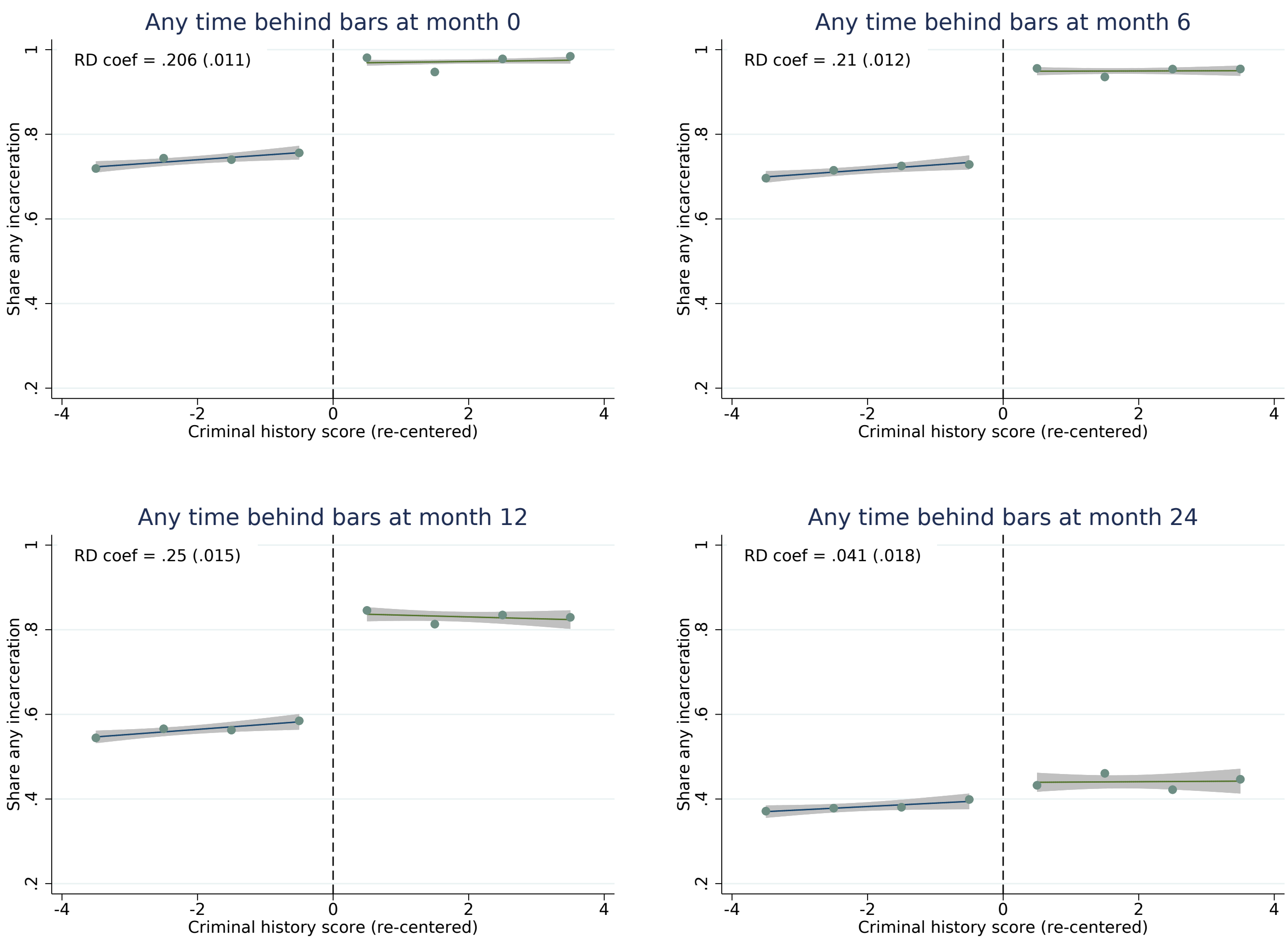
Notes: This figure shows reduced form RD estimates of being to the right of the punishment type discontinuity in class E on the likelihood of being incapacitated behind bars at month t after the date of conviction. The x-axis shows the recentered value of prior record points. The y-axis reports the share of individual who spent any time behind bars at month t after conviction. For example, the y-axis in the upper-left plot shows the share who where incarcerated for some time at month 0, which is exactly the first stage. Equivalently, the y-axis in the lower-right plot shows the share of offenders who where incarcerated for some time at month 24 after the date of conviction. Our parameter of interest, which is reported in each of the plots (i.e., RD coef), is the coefficient on an indicator for whether the individual is above the punishment type discontinuity or not. The figure includes only offenders convicted of a class F felony offense. Standard errors are clustered at the individual level.

Figure H.3: Dynamic differences in incarceration status *at* a given month after conviction (class G felony offenses)



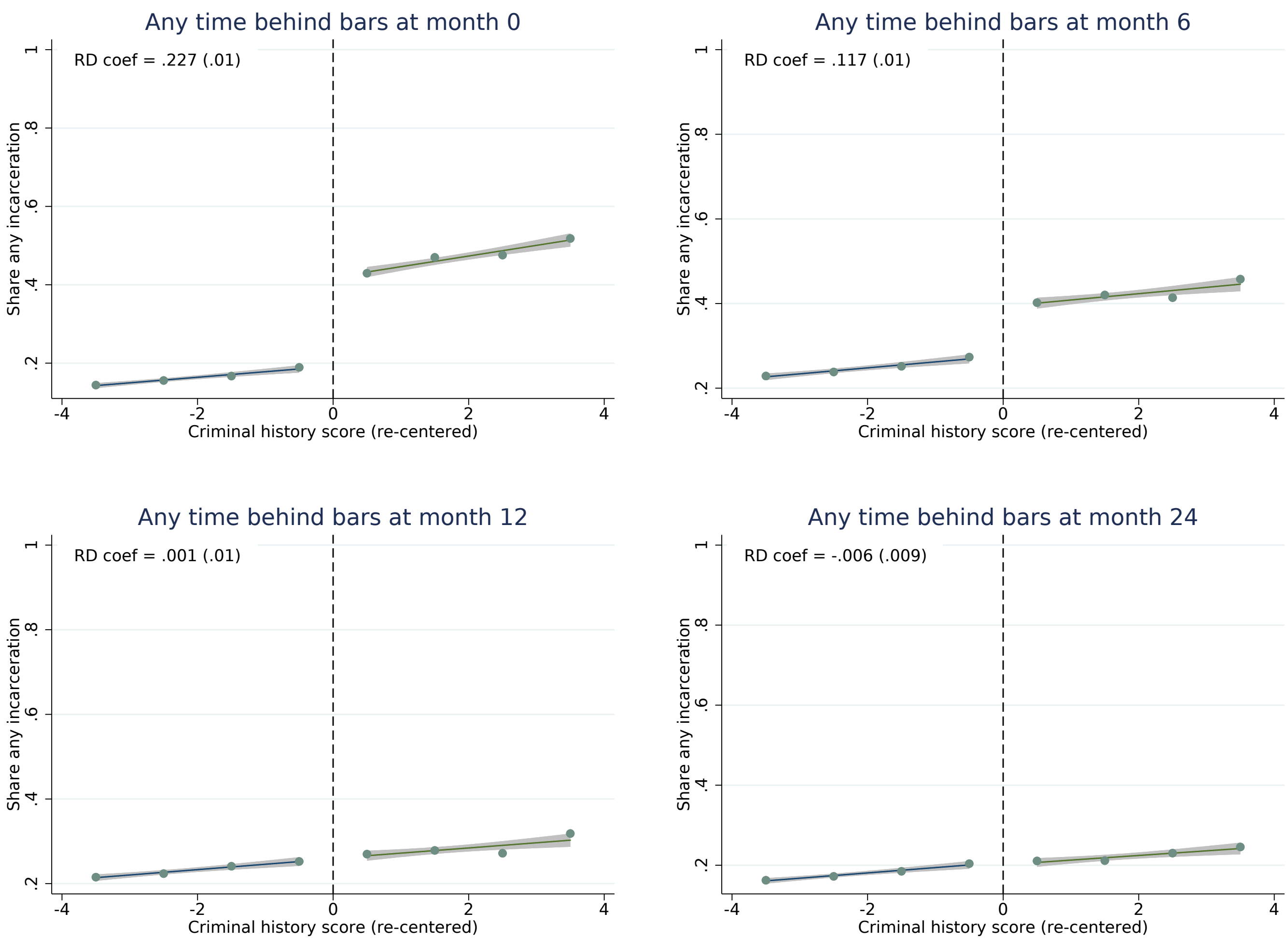
Notes: See notes to Figure H.3

Figure H.4: Dynamic differences in incarceration status *at* a given month after conviction (class H felony offenses)



Notes: See notes to Figure H.4

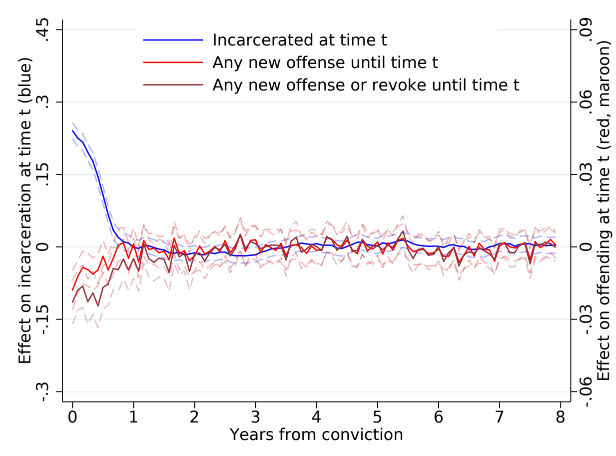
Figure H.5: Dynamic differences in incarceration status *at* a given month after conviction (class I felony offenses)



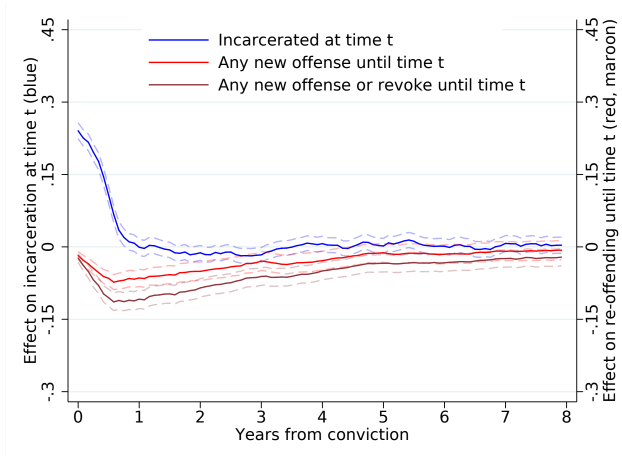
Notes: See notes to Figure H.5

Figure H.6: Reduced form estimates of reoffending *at* period t from conviction and also estimates of *any* reoffending up to period t from conviction

Class I

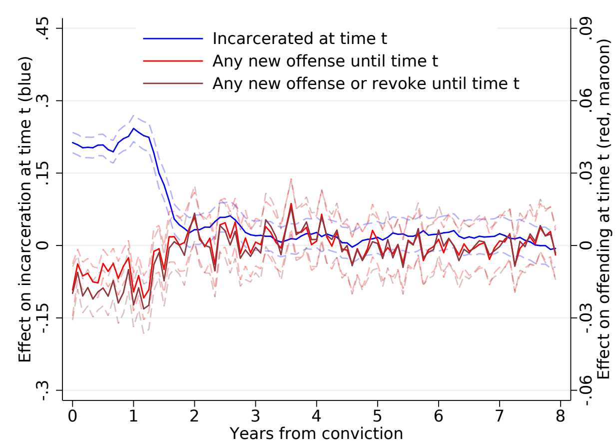


(a) reoffending at period t from conviction

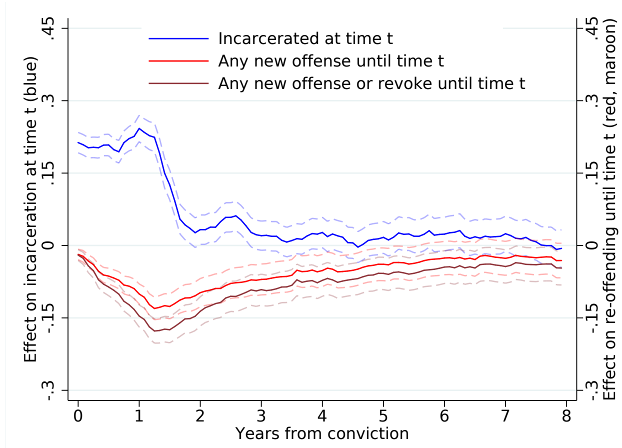


(b) Any reoffending until period t from conviction

Class H



(a) reoffending at period t from conviction

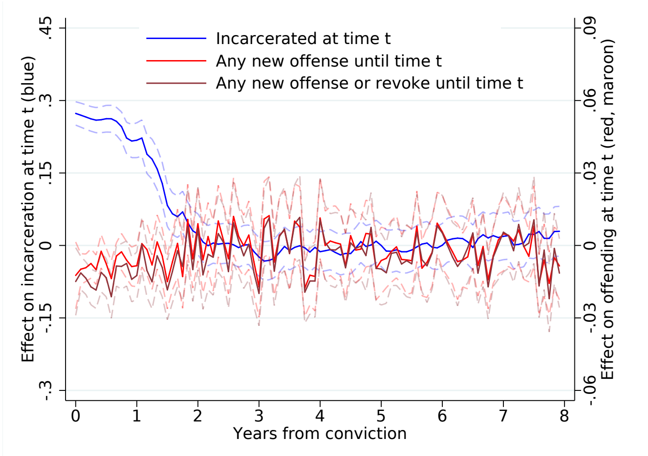


(b) Any reoffending until period t from conviction

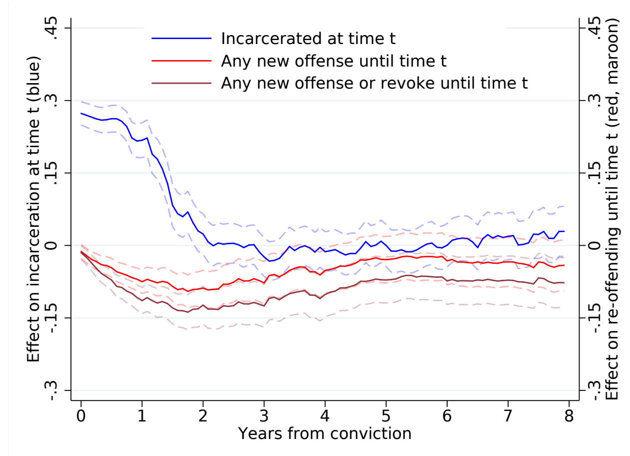
Notes: This figures shows reduced form estimates of being to the right of a punishment type discontinuity on several different outcomes of interest. All outcomes/measures are with respect to the conviction date. The blue line (left y-axis) on both panels represents the the reduced form effect on an indicator for spending any positive amount of time behind bars *at* month t from conviction. In Panel (a), the red color line (right y-axis) reports the reduced form effects on committing a new offense *at* month t, and the maroon color line (right y-axis) the estimates when also including probation volations as offending. In Panel (b), the red color line (right y-axis) reports the reduced form effects on committing *any* new offense *until* month t, and the maroon color line (right y-axis) the estimates when also including probation volations as offending. The reduced form coefficients are estimated using Equation (4), when the dependent variable is various outcomes of interest. Standard errors are clustered by individual. The regression specifications include as controls demographics (e.g., race, gender, age FEs), criminal history FEs for the duration of time previously incarcerated, the number of past incarceration spells and the number of past convictions, county FEs, and year FEs. Estimates without controls yield similar results (see for example Table 2).

Figure H.7: Reduced form estimates of reoffending *at* period t from conviction and also estimates of *any* reoffending up to period t from conviction

Class G

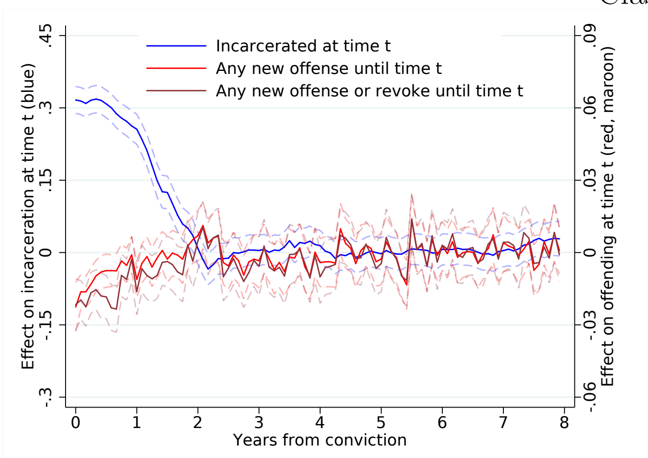


(a) reoffending at period t from conviction

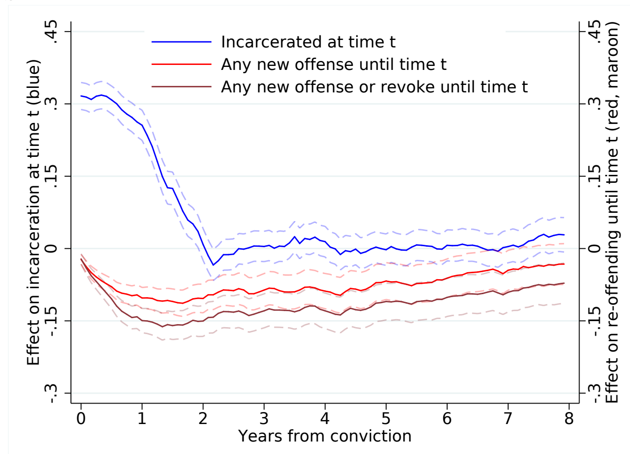


(b) Any reoffending until period t from conviction

Class F

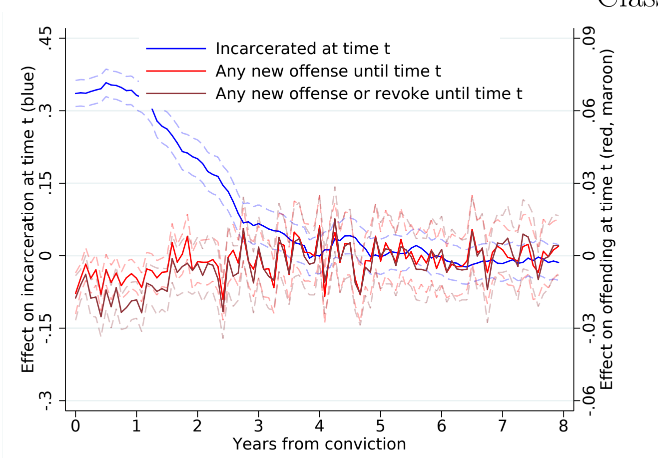


(a) reoffending at period t from conviction

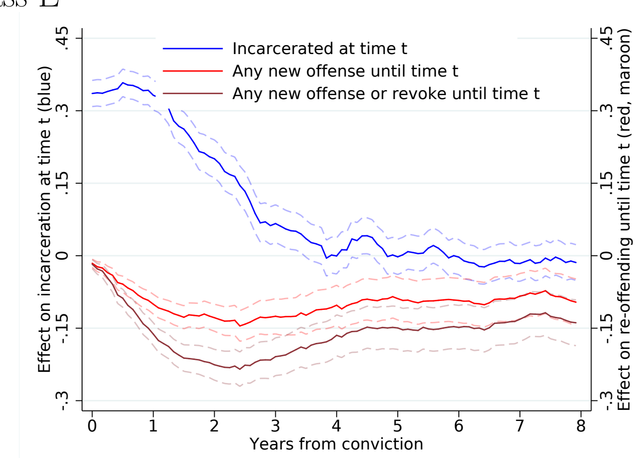


(b) Any reoffending until period t from conviction

Class E



(a) reoffending at period t from conviction



(b) Any reoffending until period t from conviction

Notes: See notes of above Figure H.6.

I Compliers characteristics

In this Appendix, we discuss how to calculate the characteristics of the compliers, i.e., the individuals who's incarceration duration is influenced by the instruments, and present estimates of the compliers characteristics.

Proposition 1. *Let* $D_i \in \{0, 1, \ldots, \bar{D}\}$ *be a discrete treatment,* $Z_i \in \{0, 1\}$ *a binary instrument, and* X_i *a pre-treatment characteristic (e.g., sex, age at offense). Assume that* $X_i \perp\!\!\!\perp Z_i$ *and all the ACR assumptions are satisfied (Angrist and Imbens, 1995), then:*

$$\frac{\mathbb{E}\left[X_i 1(D_i \geq j)|Z_i = 1\right] - \mathbb{E}\left[X_i 1(D_i \geq j)|Z_i = 0\right]}{\mathbb{E}\left[1(D_i \geq j)|Z_i = 1\right] - \mathbb{E}\left[1(D_i \geq j)|Z_i = 0\right]} = \mathbb{E}\left[X_i|D(1) \geq j > D(0)\right] \tag{I.1}$$

and

$$\frac{\mathbb{E}\left[X_i D_i|Z_i = 1\right] - \mathbb{E}\left[X_i D_i|Z_i = 0\right]}{\mathbb{E}\left[D_i|Z_i = 1\right] - \mathbb{E}\left[D_i|Z_i = 0\right]} = \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}\left[X_i|D_i(1) \geq d > D_i(0)\right] \tag{I.2}$$

where $\omega_d = \frac{Pr[D_i(1) \geq d > D_i(0)]}{\sum_{j=1}^{\bar{D}} Pr[D_i(1) \geq j > D_i(0)]}$.

Note that Proposition (1) implies that $\mathbb{E}\left[X_i|D(1) > D(0) = 0\right]$ is identifiable for $j = 1$:

$$\frac{\mathbb{E}\left[X_i 1(D_i > 0)|Z_i = 1\right] - \mathbb{E}\left[X_i 1(D_i > 0)|Z_i = 0\right]}{\mathbb{E}\left[1(D_i > 0)|Z_i = 1\right] - \mathbb{E}\left[1(D_i > 0)|Z_i = 0\right]} = \mathbb{E}\left[X_i|D(1) > D(0) = 0\right]$$

Table I.1: Compliers characteristics by felony class

	Class E		Class F		Class G		Class H		Class I	
	(1) ACR	(2) $E[X \mid D(1) > D(0) = 0]$	(3) ACR	(4) $E[X \mid D(1) > D(0) = 0]$	(5) ACR	(6) $E[X \mid D(1) > D(0) = 0]$	(7) ACR	(8) $E[X \mid D(1) > D(0) = 0]$	(9) ACR	(10) $E[X \mid D(1) > D(0) = 0]$
Black	0.736***	0.797***	0.338***	0.412***	0.510***	0.532***	0.495***	0.526***	0.625***	0.670***
	(0.0407)	(0.0516)	(0.0563)	(0.0470)	(0.0848)	(0.0669)	(0.0526)	(0.0661)	(0.0418)	(0.0209)
Age at offense	26.46***	26.92***	34.91***	35.11***	37.20***	38.69***	39.48***	39.53***	30.85***	32.73***
	(0.786)	(1.027)	(1.110)	(0.920)	(1.627)	(1.308)	(0.892)	(1.074)	(0.738)	(0.390)
Male	0.936***	0.940***	0.950***	0.967***	0.890***	0.893***	0.905***	0.905***	0.887***	0.899***
	(0.0188)	(0.0241)	(0.0226)	(0.0190)	(0.0344)	(0.0337)	(0.0233)	(0.0317)	(0.0215)	(0.0135)
Any prev. incar.	0.644***	0.640***	0.728***	0.783***	0.880***	0.869***	0.972***	0.965***	0.861***	0.863***
	(0.0390)	(0.0508)	(0.0382)	(0.0318)	(0.0312)	(0.0312)	(0.0109)	(0.0188)	(0.0209)	(0.0140)
Prev. incar. duration	9.871***	8.693***	5.327	20.87***	43.28***	39.70***	89.27***	86.01***	9.379	20.01***
	(1.688)	(2.255)	(5.010)	(3.372)	(8.627)	(6.547)	(5.259)	(6.600)	(5.192)	(1.578)
Num. prev. convictions	2.835***	2.722***	3.926***	5.107***	9.463***	9.460***	12.91***	12.96***	5.128***	5.761***
	(0.219)	(0.286)	(0.480)	(0.388)	(0.805)	(0.650)	(0.498)	(0.643)	(0.470)	(0.173)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays the mean of two types of compliers. Complier means for each characteristics are calculated as the coefficient on treatment (D_i for ACR compliers or $1(D_i > 0)$ for only extensive margin compliers) in a 2SLS regression of D_i (or $1(D_i > 0)$) multiplied by the characteristic (X_i) and using the discontinuity as the instrument (Z_i). Always-taker and never-taker means are not reported but could potentially also be calculated in an analogous 2SLS regressions of $1(D_i > 0)(1 - Z_i)X_i$ on $1(D_i > 0)(1 - Z_i)$ and $(1 - 1(D_i > 0))Z_iX_i$ on $(1 - 1(D_i > 0))Z_i$, respectively. Means for each of the five punishment type discontinuities are reported in Table I.1. Standard errors are clustered at the individual level.

J An upper bound of the incapacitation channel

We begin by showing how Equation (6) can be used to decompose reduced from estimates from conviction to an upper bound on the purely mechanical incapacitation component and a residual term that can be attributed to "behavioral effects", i.e., a residual component that cannot be accounted for only by incapacitation. The decomposition is based on two steps. The first is to note that under the null that incarceration has no impacts on criminal behavior, and its only effects are through incapacitation, the following equality follows

$$\mathbb{E}\left[Y_{i,t}(d)\right] = \mathbb{E}\left[Y_{i,t-d}(0)\right] \tag{J.1}$$

and the reduced form is simplified to

$$\begin{aligned}
&\mathbb{E}\left[Y_{i,t}|Z_i = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i = 0\right] = \\
&\quad \underbrace{\sum_{d=1}^{\bar{D}} \mathbb{E}\left[Y_{i,t-d}(0) - Y_{i,t-d+1}(0)|D_i(1) \geq d > D_i(0)\right] \Pr(D_i(1) \geq d > D_i(0))}_{\text{Only incapacitation}} + \text{“Behavioral residual”}
\end{aligned} \tag{J.2}$$

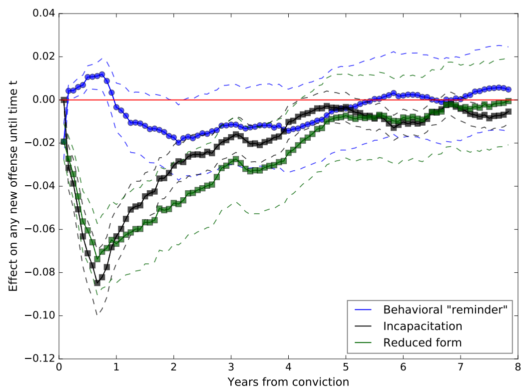
The second step that is required for the decomposition is to make an assumption on the comparability of different types of compliers. When Assumption (J.3) is satisfied, the incapacitation channel, under the null of no behavioral effects, in Equation (J.2) is identified. Assumption (J.3) imposes that individuals who have been shifted to longer terms of incarceration due to the instrument are atleast as likely to reoffend under the probation regime, i.e., if not incarcerated at all. This assumption seems plausible given Figure 7 that shows selection into incarceration based on the likelihood of reoffending and have been discussed in section 5.4.

Assumption 1.

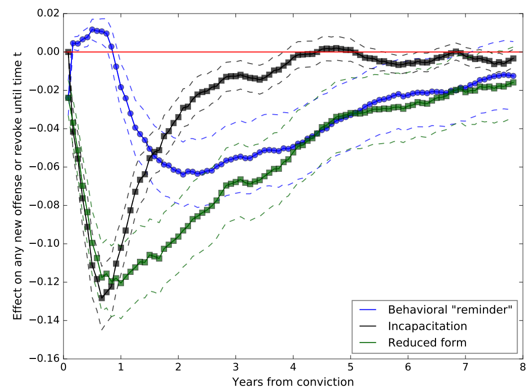
$$\mathbb{E}\left[Y_{i,t}(0)|D_i(1) \geq j > D_i(0)\right] \geq \mathbb{E}\left[Y_{i,t}(0)|D_i(1) \geq 1 > D_i(0) = 0\right] \quad \forall\, t, j \tag{J.3}$$

In Figure J.1, we decompose the reduced form estimates from offense classes E and I (most and least severe offenses) to a purely incapacitation component and a residual term that is attributed to behavioral factors. In Panels A and C, the outcome is committing any new offense within t months of conviction and in Panels B and D the outcome is any new offense or probation revocation. The estimates in Panel A (and C) show that incapacitation alone can explain almost all of the reduced form differences, suggesting that behavioral effects are negligible. However, when reoffending is defined as committing a new offense or a probation revocation (Panels B and D), then we find that incarceration has substantial crime reducing effects on criminal behavior. In Figure J.2, a similar decomposition is presented for offenses classes F, G, and H.

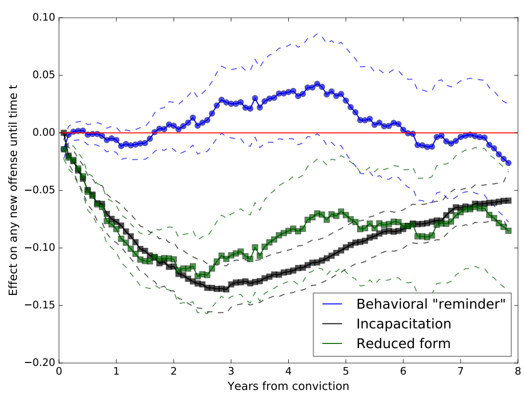
Figure J.1: Decomposition, by offense class, of reduced form estimates to a purely incapacitation component and a residaul term that is attributed to behavioral channels



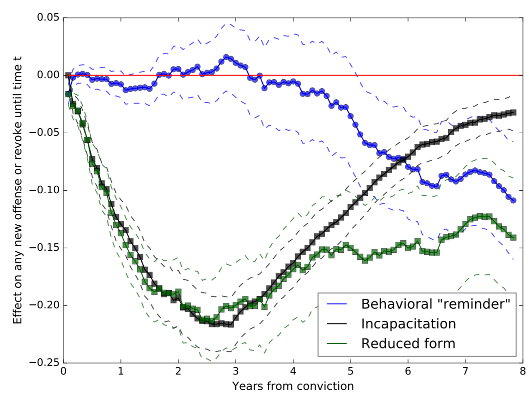
(a) Class I: Any new offense



(b) Class I: Any new offense or revoke



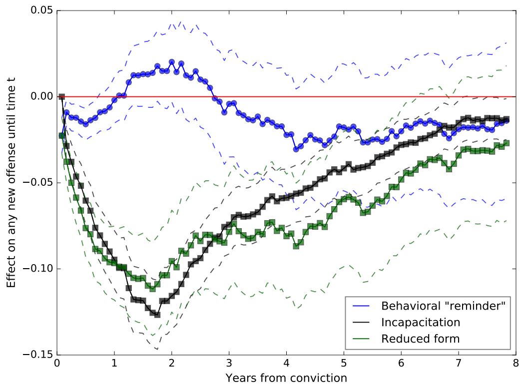
(c) Class E: Any new offense



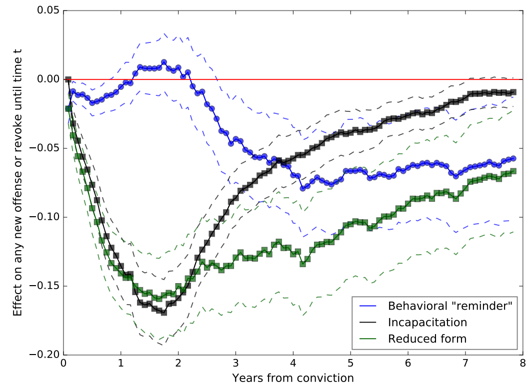
(d) Class E: Any new offense or revoke

Notes: This figure shows results of the decomposition, in Equation (J.2), of the reduced form estimates to a component that can be fully accounted for by the mechnical effect of incapacitation and a residual term that is attributed to behavioral factors. The decomposition procedure provides an upper bound on the incapacitation component and an associated lower bound on the behavioral component. The figure reports estimates for offense classes I and E, the least and most severe offenses in our data. In Panels A and C the outcome is committing any new offnse up to time t and in Panels B and D probation revocations are also classified as reoffending in addition to new offenses.

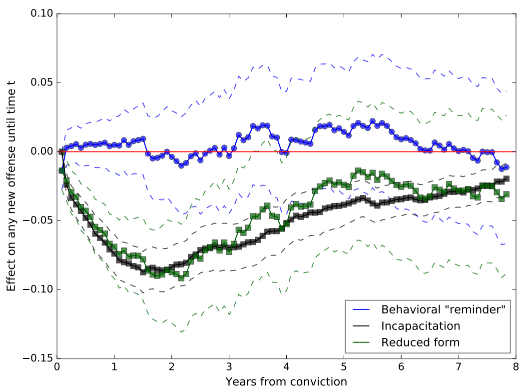
Figure J.2: Decomposition, by offense class, of reduced form estimates to a purely incapacitation component and a residaul term that is attributed to behavioral channels



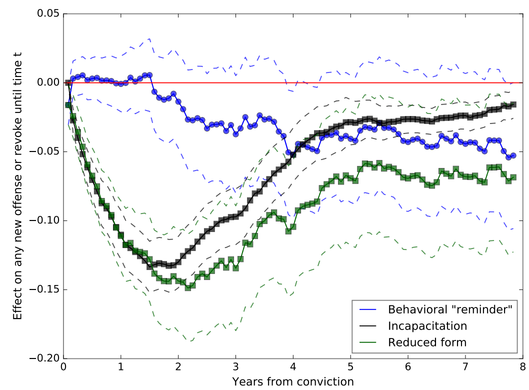
A. Class F: Any new offense



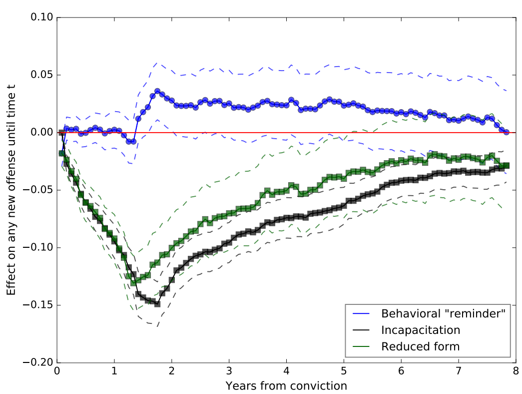
B. Class F: Any new offense or revoke



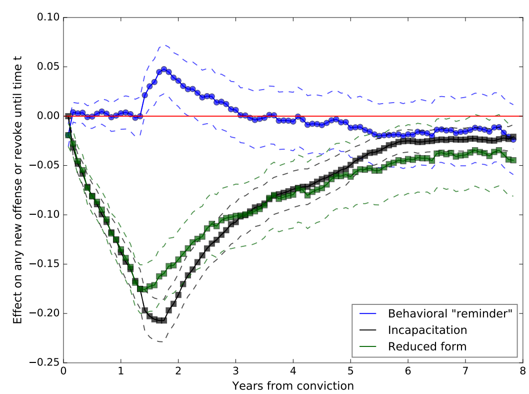
C. Class G: Any new offense



D. Class G: Any new offense or revoke



E. Class H: Any new offense



F. Class H: Any new offense or revoke

Notes: See notes to figure J.1.

K Additional robustness and validity tests

K.1 Plea bargaining does not impact our results

In this appendix, we present another way of testing that plea bargaining does not bias our estimates (in addition to the estimates in Table K.1 that are discussed in the main text). We examine whether plea bargainers are selected by taking all individuals convicted in a given offense class and prior record points value and comparing those who were initially charged in that offense class to those who plead down from more severe offenses. Since the key concern for our research design is that this type of sorting *increases* at the discontinuity, we compare these two groups of offenders just before and just after a major discontinuity.

We document that both groups also face the same punishment regime and similar exposure to incarceration. According to Appendix Figure K.1 there is no evidence that individuals initially charged with a more severe offense are incarcerated more. This result holds for both individuals just before or just after a punishment type discontinuity. Given that the two groups receive similar levels of punishment, any observed differences in reoffending should arise through selection. Appendix Figure K.2 shows that the two groups—those "Charged same felony class" and those "Charged higher felony class"—have the same likelihood of reoffending within three years after being released and also within three years from the conviction date. To conclude, we find no evidence that our results are influenced by plea bargaining.

Table K.1: Estimates of the effect of incarceration on reoffending from sentencing using charged vs. convicted offense class

	New offense			New offense of revoke			Re-incarceration		
	(1) Arraigned	(2) Charged	(3) Convicted	(4) Arraigned	(5) Charged	(6) Convicted	(7) Arraigned	(8) Charged	(9) Convicted
Months incarcerated	-0.00959***	-0.00960***	-0.00923***	-0.0146***	-0.0146***	-0.0144***	-0.0169***	-0.0168***	-0.0156***
	(0.00165)	(0.00164)	(0.00102)	(0.00166)	(0.00166)	(0.00104)	(0.00165)	(0.00164)	(0.00100)
N	363360	363360	363360	363360	363360	363360	363360	363360	363360
Dep. var. mean	0.428	0.428	0.428	0.544	0.544	0.544	0.419	0.419	0.419

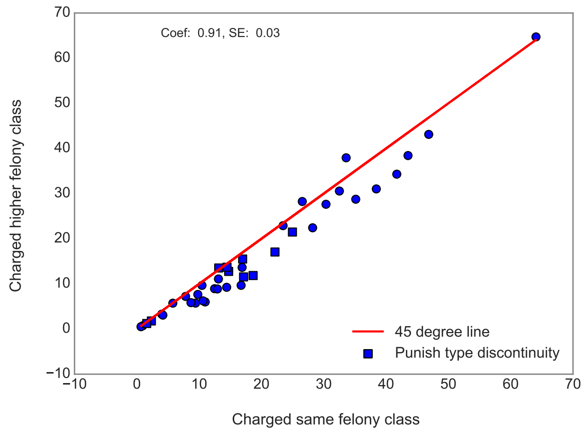
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

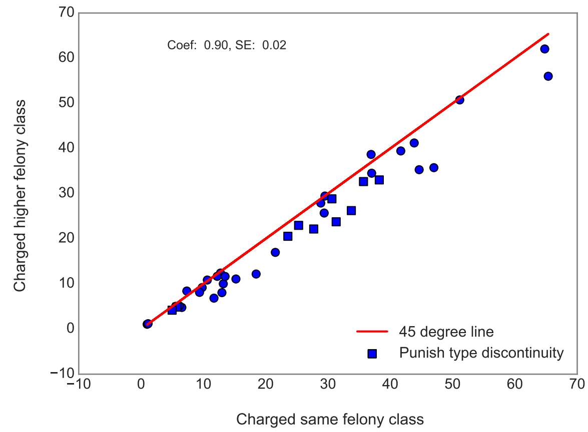
Notes: This table reports 2SLS estimates of incarceration length (D_i) on reoffending within three years of conviction according to three different measures of reoffending. For each measure of reoffending (e.g., New offense), three estimates are reported. Each column shows the estimated effect when calculating the instruments using a different classification of offenses felony severity classes. The first column uses the offense that the individual was arrested for, The second column the offense that she was arraigned for, and lastly the third column the offense she got convicted of. In our main analysis we use the third column. It is clear that the estimates in all columns are similar, however, the standard errors in the third column are substantially lower. Standard errors are clustered at the individual level.

Figure K.1: Difference in punishment between plead down and same charged offender

Months incarcerated

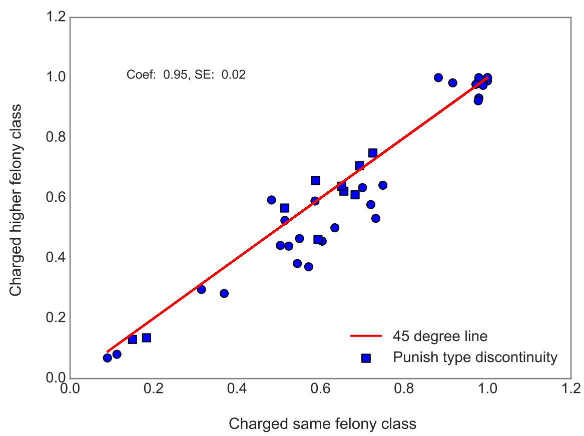


(a) Just *before* discontinuity

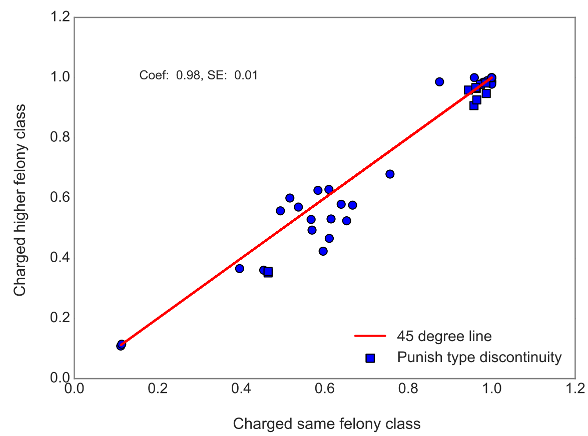


(b) Just *after* discontinuity

Share incarcerated



(a) Just *before* discontinuity

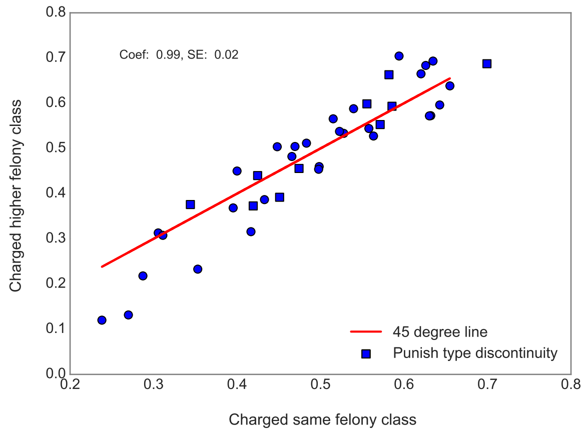


(b) Just *after* discontinuity

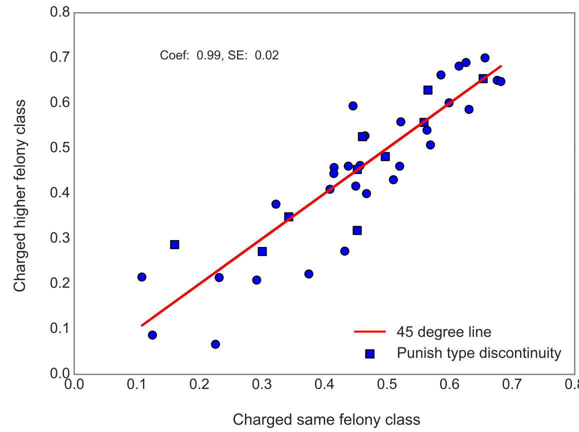
Notes: See the notes in Figure K.2.

Figure K.2: Reoffending rates between plead down and same charged offenders

Reoffending estimates from release (at-risk)

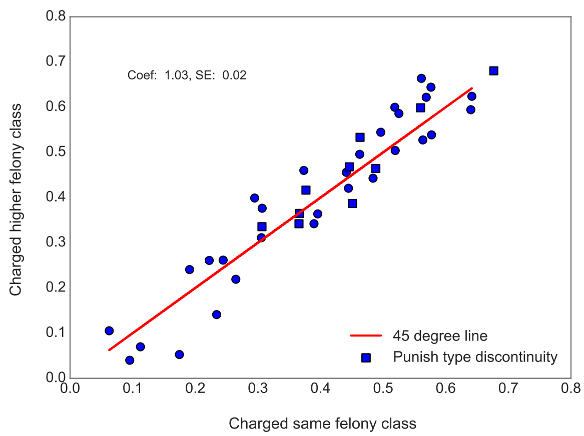


(a) Just *before* discontinuity

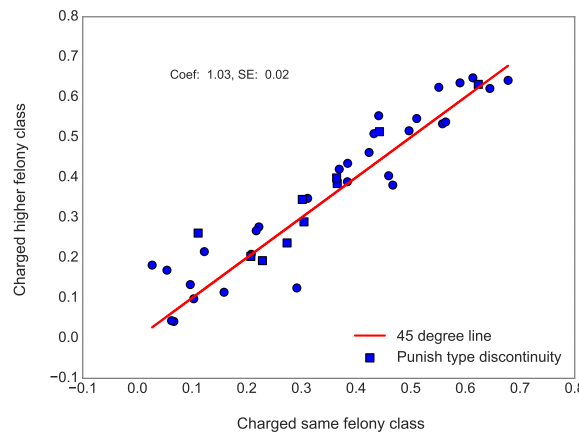


(b) Just *after* discontinuity

Reoffending estimates from conviction



(c) Just *before* discontinuity



(d) Just *after* discontinuity

Notes: This figure splits all individuals convicted in a given offense class and prior record points value and compares those who were initially charged in that offense class (x-axis) to those who plead down from more severe offenses (y-axis). Since the key concern for our research design is that this type of sorting increases at the discontinuity, we compare these two groups of offenders just before (left panel plots) and just after (right panel plots) a major discontinuity.

K.2 No evidence of differences in detection

The tables and figures below are discussed in Section 6 in the main text.

Table K.2: 2SLS estimates of the effect of incarceration length on re-offending within 3 years using different parts of the grid

	Re-incarceration		Any new offene		Felony		Assault	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ext.	Int.	Ext.	Int.	Ext.	Int.	Ext.	Int.
Months incap	-0.0138***	-0.0189***	-0.00793***	-0.0206***	-0.00600***	-0.0200***	-0.00217***	-0.00312*
	(0.000724)	(0.00226)	(0.000751)	(0.00256)	(0.000702)	(0.00250)	(0.000435)	(0.00125)
N	491135	491135	491135	491135	491135	491135	491135	491135
Dep. var. mean	0.401	0.401	0.417	0.417	0.306	0.306	0.0706	0.0706
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	209.2	11.96	209.2	11.96	209.2	11.96	209.2	11.96
J stat	58.32	39.61	2.672	29.42	6.021	39.15	2.471	10.02
J stat p	6.54e-12	0.000293	0.614	0.00915	0.198	0.000346	0.650	0.761
Hausman p	3.37e-12	0.0000754	0.000927	3.13e-10	0.00274	7.03e-11	0.00574	0.187

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports 2SLS estimates of incarceration length (D_i) on reoffending within three years of conviction according to four different measures of reoffending. For each measure of reoffending (e.g., Any new offense), two estimates are reported. Each column shows the estimated effect when using a different set of discontinuities as the excluded-instruments. The first column (Ext.) uses the five punishment type discontinuities as the excluded-instruments. The second column (Int.) uses only the 15 discontinuities that shift only the intensive margin of the length of incarceration and do not impact the type of punishment (incarceration vs. probation). Standard errors are clustered at the individual level.

Table K.3: Independent risks: 2SLS estimates of the effect of incarceration length on re-offending within 3 years using different parts of the grid

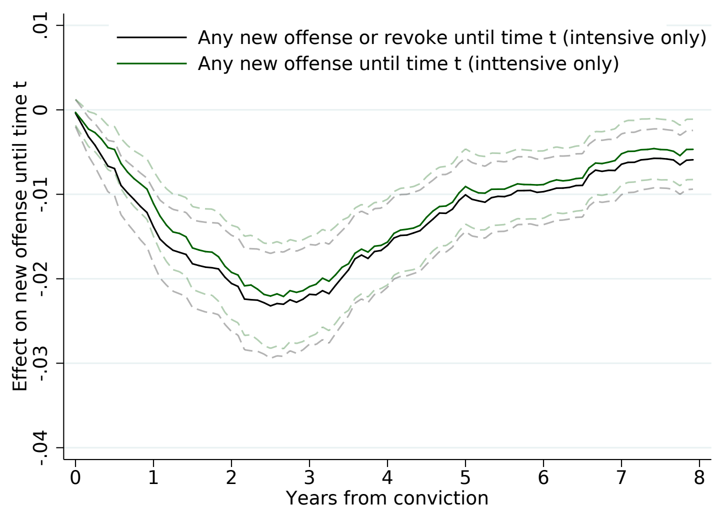
	Re-incarceration		Any new offene		Felony		Assault	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ext.	Int.	Ext.	Int.	Ext.	Int.	Ext.	Int.
Months incap	-0.00995***	-0.0168***	-0.0105***	-0.0202***	-0.00774***	-0.0198***	-0.00286***	-0.00334*
	(0.000773)	(0.00221)	(0.000860)	(0.00255)	(0.000807)	(0.00253)	(0.000502)	(0.00130)
N	411246	411246	411246	411246	411246	411246	411246	411246
Dep. var. mean	0.285	0.285	0.434	0.434	0.321	0.321	0.0734	0.0734
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	160.6	11.05	160.6	11.05	160.6	11.05	160.6	11.05
J stat	31.78	36.49	6.065	29.04	7.166	39.38	2.668	8.928
J stat p	0.00000212	0.000882	0.194	0.0103	0.127	0.000318	0.615	0.836
Hausman p	0.00000795	0.0000104	0.00000174	9.04e-09	0.000137	1.27e-09	0.000604	0.180

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

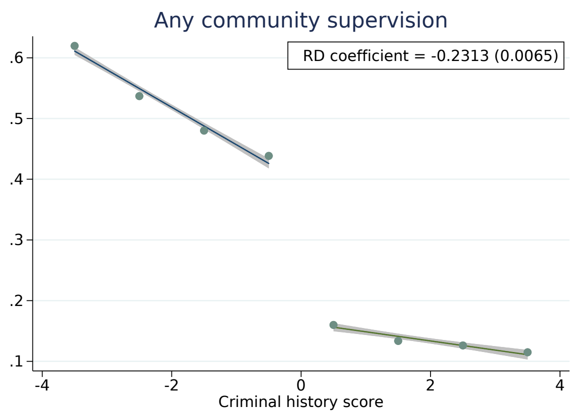
Notes: See notes of Table K.2. The only different between the two tables is that in this table we used an independent risks assumption and dropped from the sample all offenders with a probation revocation prior to a new offense within three years of conviction.

Figure K.3: The effect of length of incarceration on re-offending using only intensive margin variation



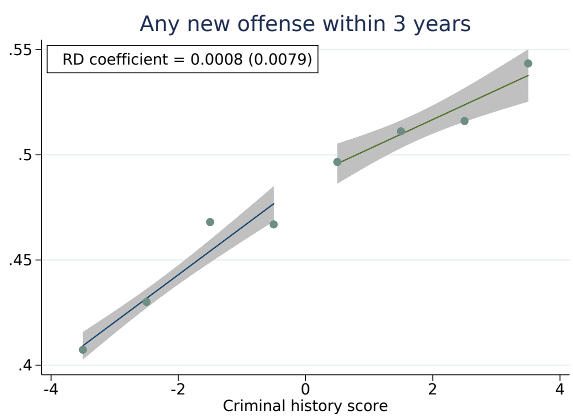
Notes: This figure reports 2SLS estimates of incarceration length (D_i) on reoffending within t months from the sentencing date. Two measures of reoffending are used. The first is an indicator for whether the individual committed any new offense until month t from sentencing (green line). The second includes also probation revocations in the reoffending indicator. All estimates are from a 2SLS that uses *only* the 15 discontinuities that shift only the intensive margin of the length of incarceration and do not impact the type of punishment (incarceration vs. probation). Standard errors are clustered at the individual level.

Figure K.4: Effects on the type of punishment (community vs. intermediate supervision) and future re-offending and re-incarceration within 3 years of conviction

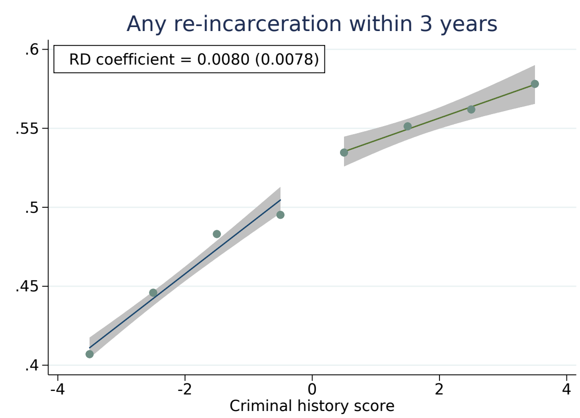


(a) Community supervision

(b) Intermediate supervision



(c) Any new offense 3 years from conviction

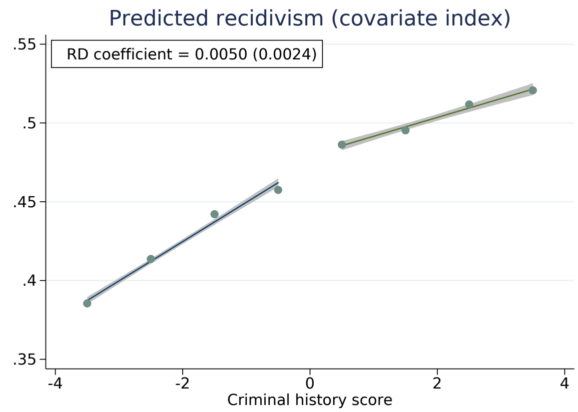


(d) Any re-incarceration 3 years from conviction

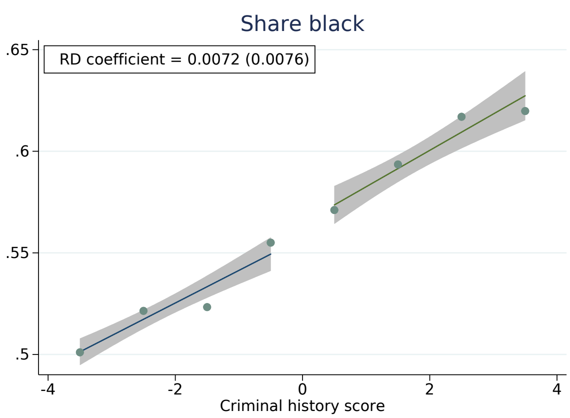
Notes: This figure shows the impacts of the discontinuity in the type of probation supervision (community vs. intermediate) in felony offense class I, when moving between prior record levels II and III, on the type of probation supervision . The plots in the first row show that the transition between prior record levels has a sailiant effect on the type of supervision that offenders are assigned. The plots in the second row show that the discontinuity does not have an influence on re-offending outcomes such as committing a new offense or being re-incarcerated within three years of the time of conviction.

Figure K.5: Validity checks that incarceration exposure and pre-conviction controls vary smoothly at discontinuity

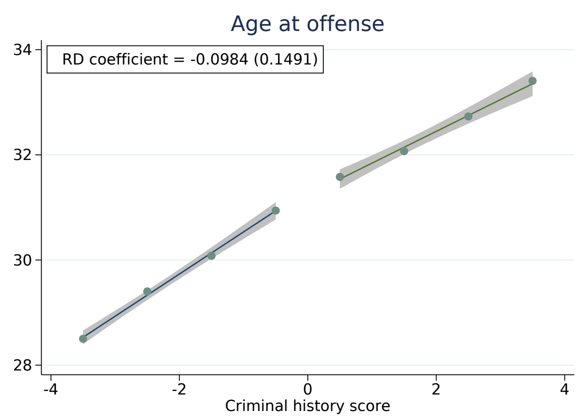
(a) Any incarceration sentence



(b) Predicted likelihood of committing a new offense



(c) Share black



(d) Age at offense

Notes: This figure shows the impacts of the discontinuity in the type of probation supervision (community vs. intermediate) in felony offense class I, when moving between prior record levels II and III, on outcomes that are not supposed to be influenced by the discontinuity. This figure presents validity checks that support a causal interpretation to the estimates effects in Figure K.4.

L Effects of incarceration on reoffending from release

In this appendix, we discuss the main reduced form effects of the punishment discontinuities on reoffending post-release. The main results are summarized graphically in Figure L.1. Panel (a), shows the effects on the likelihood of being behind bars (blue line), committing a new offense (red line) and committing a new offense or a probation revocation (maroon line) *at* period t. The punishment type discontinuities cause a large *negative* effect on incarceration at month t from release, as is expected given the impact of revocations in the untreated group. While the effects decline steadily over the following months, it takes four years until the instruments are no longer predictive of incarceration status.

The red and maroon lines in Panel (a) plot differences in the likelihood of committing a new offense (and probation revocation) *at* period t from at-risk. These estimates are noisily centered around zero and show no systematic differences, despite the fact that the control group has higher likelihoods of being incapacitated behind bars due to probation revocations. When including probation revocations as reoffending (maroon line), there are decreases in the likelihood of reoffending in the first 2-3 years.

Panel (b) shows the effects on the likelihood of *any* reoffending from at-risk to period t. The estimates are negative but not statistically distinguishable from zero. This is somewhat surprising given the differences in incarceration rates over this period. However, if probation revocations are also included as reoffending, the estimates show large reductions in crime among the initially "treated" population.

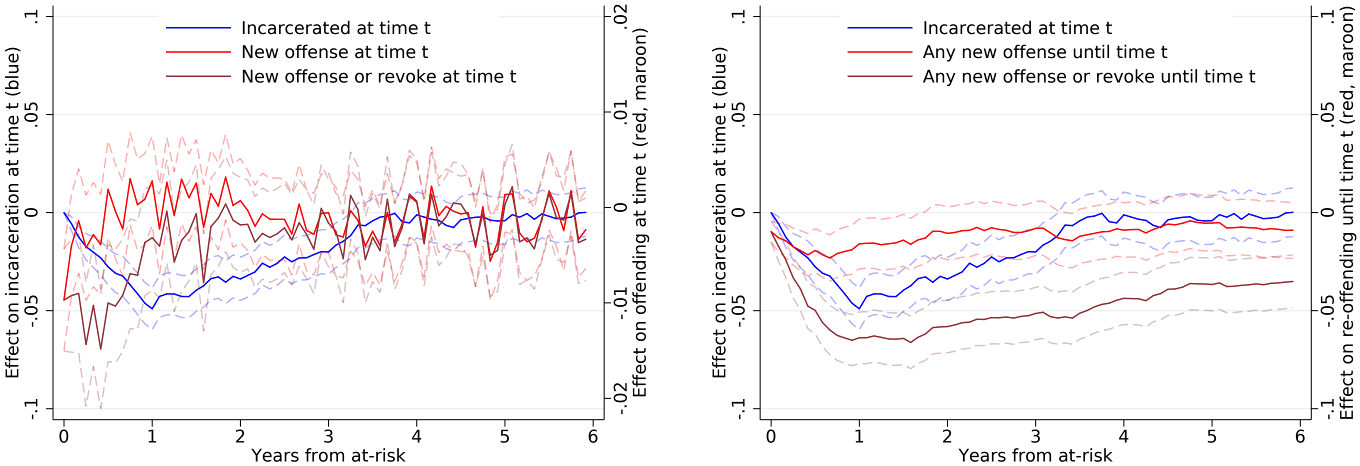
Table L.1: Independent risks: Estimates of incarceration effects on different types of new offenses using measures from at-risk

	Measure of crime					
	(1)	(2)	(3)	(4)	(5)	(6)
	Re-incarceration	Any new offense	Felony	Assault	Property	Drug
Months incap	-0.00328***	-0.00307**	-0.00124	-0.00109	0.000509	0.000354
	(0.000970)	(0.000986)	(0.000979)	(0.000704)	(0.000829)	(0.000790)
N	397060	397060	397060	397060	397060	397060
Dep. var. mean among non-incarcerated	0.314	0.451	0.326	0.0725	0.172	0.178
One year effect in percentages	-12.5	-8.17	-4.55	-18.0	3.55	2.39
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-instruments)	204.2	204.2	204.2	204.2	204.2	204.2
J stat	21.09	11.39	7.136	1.559	9.020	8.947
J stat p	0.000304	0.0225	0.129	0.816	0.0606	0.0624
Hausman p	0.986	0.295	0.522	0.216	0.930	0.216
Lochner-Moretti stat	-0.0000122	-0.00172	-0.00155	-0.00103	-0.000279	0.000594
Lochner-Moretti p	0.990	0.0865	0.118	0.141	0.738	0.455

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Dependent variable is an indicator for any new offense recorded in DPS or AOC data between 0 and three years of the individual's release date. Observation in which a probation revocation occurred prior to a new offense have been dropped according to the independent risks assumption. Standard errors are clustered by individual. Standard errors are clustered by individual. Each row reports results for a different group of offenders. In each cell the first line reports the 2SLS coefficient, the second the standard error and the third the first-stage F-statistic.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure L.1: Reduced form estimates of re-offending *at* period t from at-risk and also estimates of *any* re-offending up to period t from at-risk



(a) Re-offending at period t from at-risk

(b) Any re-offending until period t from at-risk

Notes: This figure shows reduced form estimates of being to the right of a punishment type discontinuity on several different outcomes of interest. All outcomes/measures are with respect to the at-risk date which is the release date for incarcerated individuals and the conviction date for non-incarcerated individuals. The blue line (left y-axis) on both panels represents the the reduced form effect on an indicator for spending any positive amount of time behind bars *at* month t from at-risk. In Panel (a), the red color line (right y-axis) reports the reduced form effects on committing a new offense *at* month t, and the maroon color line (right y-axis) the estimates when also including probation revocations as offending. In Panel (b), the red color line (right y-axis) reports the reduced form effects on committing *any* new offense *until* month t, and the maroon color line (right y-axis) the estimates when also including probation revocations as offending. Standard errors are clustered by individual. See also the notes in Figure 6 for further details on the estimation.

M Selection Model and Control Function Approach

This appendix provides additional details, results, and extensions to the control function approach that is laid out in Section 7.

M.1 Estimation of ordered choice model

The selection model described in Equations (M.1) and (M.2) is estimated via maximum likelihood.

$$D_i = d \quad \text{if} \quad \mathbb{I}\left\{ \underbrace{C^l_{d-1}(Z^l_i)}_{\text{cut-offs}} \leq X'_i \gamma^l_0 + \underbrace{\nu_i}_{\text{Unobserved heterogeneity}} < C^l_d(\underbrace{Z^l_i}_{\text{Instrument}}) \right\} \tag{M.1}$$

where $\nu_i \sim N(0,1)$ and $l \in \{E, F, G, H, I\}$ is the class of offender i's conviction and the thresholds are weakly increasing

$$\begin{aligned} C^l_{d-1}(Z^l_i) &\leq C_d(Z^l_i) \quad \forall Z^l_i,\ l \\ C^l_{-1}(Z^l_i) = -\infty, \quad C^l_{\bar{D}}(Z^l_i) &= \infty \quad \forall Z^l_i,\ l \end{aligned} \tag{M.2}$$

To guarantee that the constraint in Equation (M.2) are satisfied, certain structure is usually imposed on the thresholds, $C_d^l(Z_i^l)$ (Greene and Hensher, 2010). In the estimation procedure, we model the cut-off values as

$$\begin{aligned} \mu_{i0}^l &\equiv Z_i^l \gamma_{10}^l + \alpha_0^{l(i)} \\ \mu_{is}^l &\equiv \exp(Z_i^l \gamma_{1d} + \alpha_d^l) \forall\ d > 0 \\ C_d^l(Z_i^l) &\equiv \sum_{l=0}^{d} \mu_{id}^l \quad \forall\ d < \bar{D} \end{aligned} \tag{M.3}$$

Note that the γ_{1d}^l parameters are identified from the variation in Z_i^l across individuals.[43]

The likelihood function is

$$L(D_1, \ldots, D_N | \gamma_0^l, \gamma_1^d, \alpha_d) = \prod_{i=1}^{N} \prod_{d=0}^{\bar{D}} 1(D_i = d) \left[\Phi(C_d(Z_i^{l(i)}) - X_i' \gamma_0^{l(i)}) - \Phi(C_{d-1}(Z_i^{l(i)}) - X_i' \gamma_0^{l(i)}) \right] \tag{M.4}$$

The associated log-likelihood is:

$$\sum_{i=1}^{N} \sum_{d=0}^{\bar{D}} 1(D_i = d) \log \left[\Phi \left(\sum_{l=0}^{d} \exp\left(Z_i \gamma_1^l + \alpha_l \right) - X_i' \gamma_0 \right) - \Phi \left(\sum_{l=0}^{d-1} \exp\left(Z_i \gamma_1^l + \alpha_l \right) - X_i' \gamma_0 \right) \right] \tag{M.5}$$

and the F.O.Cs are:

$$\frac{\partial l(\cdot)}{\partial \gamma_0} = \sum_{i=1}^{N} 1(D_i = d) \frac{1}{\Phi \left(\sum_{j=0}^{d} \exp\left(Z_i \gamma_1^j + \alpha_j \right) - X_i' \gamma_0 \right) - \Phi \left(\sum_{l=0}^{d-1} \exp\left(Z_i \gamma_1^l + \alpha_l \right) - X_i' \gamma_0 \right)} \tag{M.6}$$

$$\left[\phi \left(\sum_{l=0}^{d} \exp\left(Z_i \gamma_1^l + \alpha_l \right) - X_i' \gamma_0 \right) - \phi \left(\sum_{l=0}^{d-1} \exp\left(Z_i \gamma_1^l + \alpha_l \right) - X_i' \gamma_0 \right) \right] \cdot (-X_i')$$

[43] Notice that if $\mu_{i0} = \exp(Z_i \gamma_1^0 + \alpha_0)$, then we would have been imposing an additional constraint that all the thresholds, $C_d(Z_i)$, are strictly grater than zero. Instead we use $\mu_{i0}^{l(i)} = Z_i^{l(i)} \gamma_{10}^{l(i)} + \alpha_0^{l(i)}$ that does not impose any such restrictions.

$$\frac{\partial l(\cdot)}{\partial \alpha_d} = \sum_{i=1}^{N} 1(D_i \geq d) \frac{1}{\Phi\left(\sum_{l=0}^{D_i} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right) - \Phi\left(\sum_{l=0}^{D_i-1} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right)} \tag{M.7}$$

$$\cdot \left[\phi\left(\sum_{l=0}^{D_i} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right) - 1(D_i > d) \cdot \phi\left(\sum_{l=0}^{D_i-1} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right)\right] \cdot \exp\left(Z_i\gamma_1^d + \alpha_d\right)$$

$$\frac{\partial l(\cdot)}{\partial \gamma_1^d} = \sum_{i=1}^{N} 1(D_i \geq d) \frac{1}{\Phi\left(\sum_{l=0}^{D_i} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right) - \Phi\left(\sum_{l=0}^{D_i-1} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right)} \tag{M.8}$$

$$\cdot \left[\phi\left(\sum_{l=0}^{D_i} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right) - 1(D_i > d) \cdot \phi\left(\sum_{l=0}^{D_i-1} \exp\left(Z_i\gamma_1^l + \alpha_l\right) - X_i'\gamma_0\right)\right] \cdot \exp\left(Z_i\gamma_1^d + \alpha_d\right) Z_i'$$

M.2 Goodness of fit tests

This section of the appendix describes tests for model fit. Overall, the model fits the data well, as shown by Appendix Figure M.1. This figure bins observations into groups by felony class and prior record points. Within each group we calculate the average observed incarceration length and the average predicted length according to the ordered choice model (weighting each duration by the predicted probability of assignment for each observation). Panel (a) reports the results when using all the data to fit the ordered choice model and to assess its accuracy. In Panel (b), we randomly split the data into two parts, fit the model on one half, and conduct the accuracy comparison on the other half. The figure clearly shows that there is no over-fitting problem in our case. Both Panel (a) and (b) show similar results.

To test the model's fit in more detail, we use the model to replicate the experimental variation induced by the instruments, i.e., the ACR weights. The ordered choice model yields values of $\Pr(D_i(1) \geq d > D_i(0))$ for every instrument and incarceration level d:

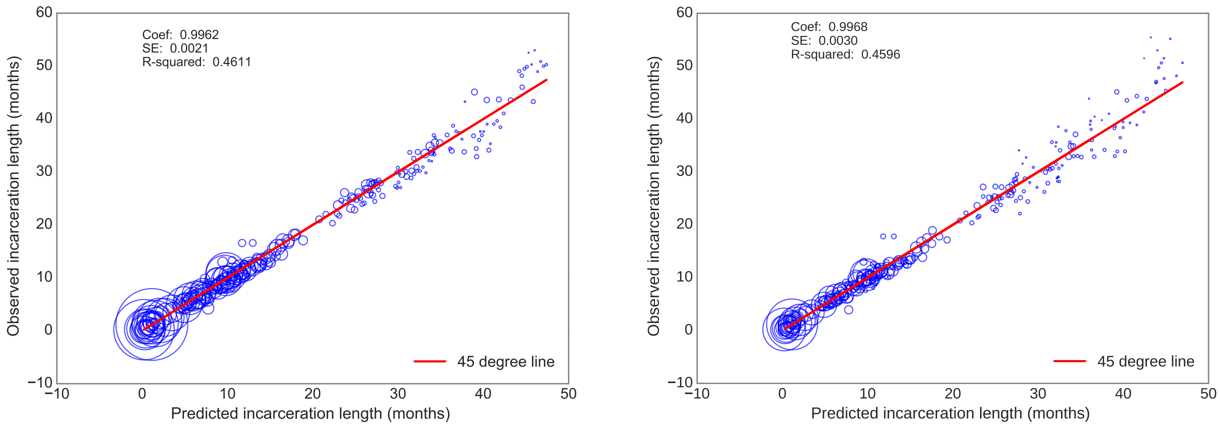
$$\begin{aligned}
\Pr(D_i(1) \geq d > D_i(0)) &= \Pr(C_{d-1}^{l(i)}(Z_i^{l(i)} = 1) - X_i\gamma_0^{l(i)} < \nu_i \leq C_{d-1}^{l(i)}(Z_i^{l(i)} = 0) - X_i'\beta_0) \\
&= \Phi\left(C_{d-1}^{l(i)}(Z_i^{l(i)} = 0) - X_i\gamma_0^{l(i)}\right) - \Phi\left(C_d^{l(i)}(Z_i^{l(i)} = 1) - X_i'\gamma_0^{l(i)}\right)
\end{aligned} \tag{M.9}$$

The same probabilities $\Pr(D_i(1) \geq d > D_i(0))$ can also be recovered non-parametrically using:

$$\Pr(D_i(1) \geq d > D_i(0)) = \mathbb{E}\left[\mathbb{1}\left\{D_i \geq d\right\} | Z_i^l = 1\right] - \mathbb{E}\left[\mathbb{1}\left\{D_i \geq d\right\} | Z_i^l = 0\right] \tag{M.10}$$

Appendix Figure M.2 shows that the non-parametric estimates of $\Pr(D_i(1) \geq d > D_i(0))$ and the model-based predictions of the same probabilities closely follow the same patterns in each class.

Figure M.1: Assessing the fit of the ordered-choice model

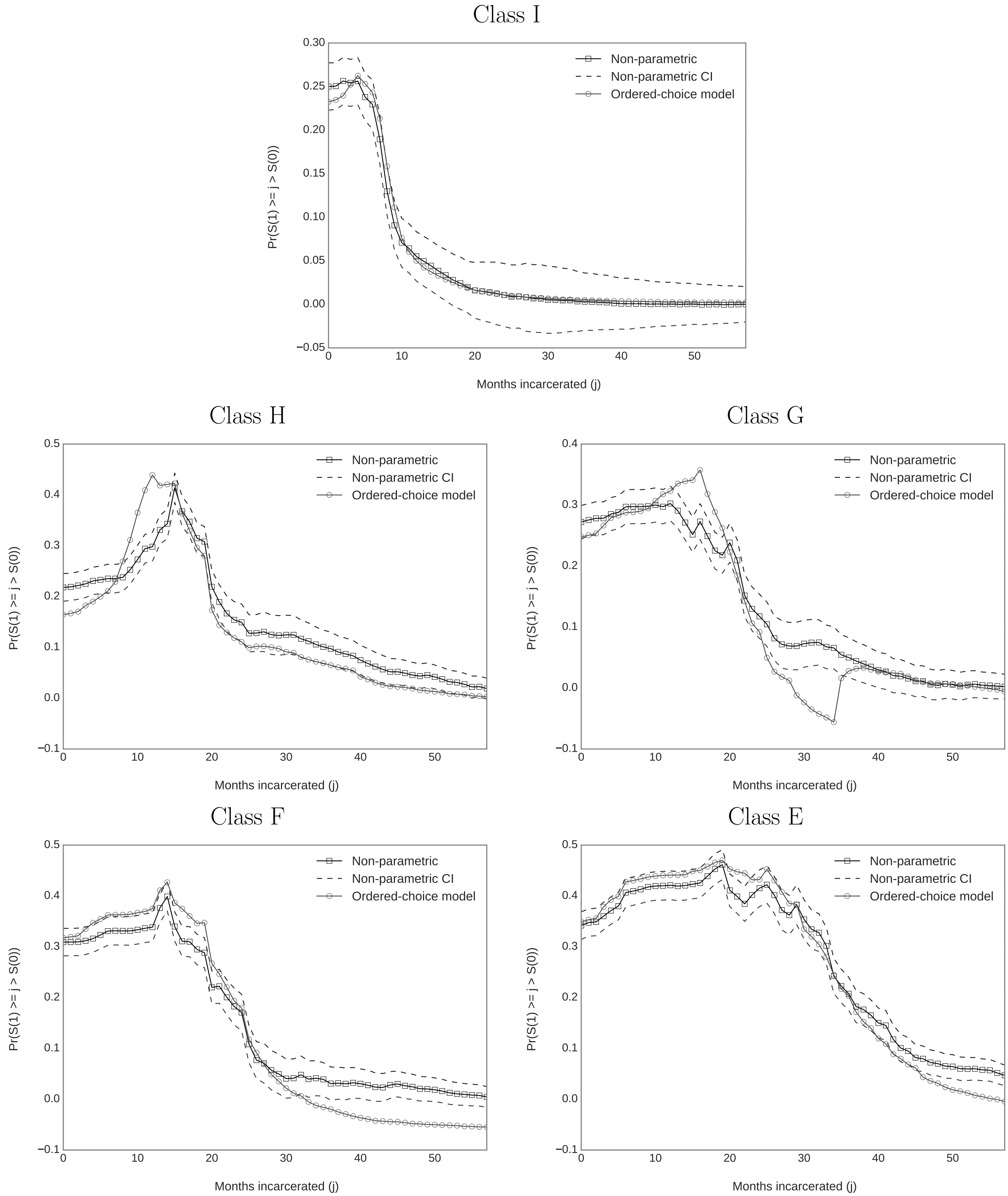


(a) All data

(b) Split sample

Notes: The figure reports the average actual (and predicted) incarceration length. Observations are binned into groups by felony class and prior record points, generating 125 points with a varying number of observations in each cell. Each point in the figure reports the average actual and predicted incarceration length in each cell. The size of the dots represents the number of observations in the cell. Cells with more observations will have larger circles. The x-axis shows the average predicted incarceration length (months) and the y-axis the actual average incarceration length (months). The red line is the 45 degree line. If the dots are below the red line then the average prediction in those cells is higher than the average observed incarceration length in the cell.

Figure M.2: Ordered-choice model replication of ACR weights across punishment type discontinuities



Notes: The x-axis report the d value and the y-axis the estimate, and confidence interval, of the following probability $\Pr[D_i(1) \geq d > D_i(0)]$. This probability can be interpreted as the probability that an individual will be shifted by the instrument from an incarceration exposure that is strictly lower than d to one which is d or higher.

M.3 Control functions

The selection correction $\lambda\left(X_i, Z_i^{l(i)}, d\right)$ is

$$\begin{aligned}\lambda\left(X_i, Z_i^{l(i)}, d\right) &\equiv \mathbb{E}\left[\nu_i | D_i = d, Z_i^l . X_i\right] \\ &= \mathbb{E}\left[\nu_i | C_{d-1}^{l(i)}(Z_i^{l(i)}) - X_i \gamma_0^{l(i)} \leq \nu_i < C_d^{l(i)}(Z_i^{l(i)}) - X_i' \gamma_0^{l(i)}\right] \\ &= \frac{\phi(C_{d-1}^{l(i)}(Z_i^{l(i)}) - X_i'\gamma_0^{l(i)}) - \phi(C_d^{l(i)}(Z_i^{l(i)}) - X_i'\gamma_0^{l(i)})}{\Phi(C_d^{l(i)}(Z_i^{l(i)}) - X_i'\gamma_0^{l(i)}) - \Phi(C_{d-1}^{l(i)}(Z_i^{l(i)}) - X_i'\gamma_0^{l(i)})}\end{aligned} \tag{M.11}$$

The selection correction $\mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0)\right]$ is a variant of the selection correction from Equation (M.11), when ν_i is restricted to the values of the compliers population of each treatment effect increment

$$\begin{aligned}\mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0), Z_i^l, X_i\right] &= \mathbb{E}\left[\nu_i | C_{d-1}^{l(i)}(Z_i^{l(i)} = 1) - X_i \gamma_0^{l(i)} \leq \nu_i < C_d^{l(i)}(Z_i^{l(i)} = 0) - X_i' \gamma_0^{l(i)}\right] \\ &= \frac{\phi(C_{d-1}^{l(i)}(Z_i^{l(i)} = 1) - X_i'\gamma_0^{l(i)}) - \phi(C_d^{l(i)}(Z_i^{l(i)} = 0) - X_i'\gamma_0^{l(i)})}{\Phi(C_d^{l(i)}(Z_i^{l(i)} = 0) - X_i'\gamma_0^{l(i)}) - \Phi(C_{d-1}^{l(i)}(Z_i^{l(i)} = 1) - X_i'\gamma_0^{l(i)})}\end{aligned} \tag{M.12}$$

M.4 Decomposition of the effect of a one month increase in incarceration on reoffending to behavioral and incapacitation channels

Another tractable decomposition of the reduced form effects from conviction to incapacitation and behavioral channels can be performed by decomposing the behavioral effects to those attributed to a change in exposure to incarceration holding time at-risk fixed ($\theta^0_{d,t-d} - \theta^0_{d-1,t-d} + \nu_i\left[\theta^1_{d,t-d} - \theta^1_{d-1,t-d}\right]$) and those due to a changes in at-risk time while holding fixed incarceration exposure at a given level ($\theta^0_{d-1,t-d} - \theta^0_{d-1,t-d-1} + \nu_i\left[\theta^1_{d-1,t-d} - \theta^1_{d-1,t-d-1}\right]$). Consider the following replication of a change in one month of exposure to incarceration (d vs. $d-1$) while holding fixed the time-varying covariates $W = w$ and using the $X_i = x$ ($\equiv \mathbb{E}\left[X_i | D_i(1) \geq d > D_i(0)\right]$) and

$\nu_i = \mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0)\right]$ characteristics of compliers (holding fixed $D_i = d-1$)

$$\begin{aligned}
&\mathbb{E}\left[Y_{i,t}(d)|X_i, Z_i, W_{i,d}, \nu_i\right] - \mathbb{E}\left[Y_{i,t}(d-1)|X_i, Z_i, W_{i,d-1}, \nu_i\right] = \\
&\underbrace{X_i'(\xi_{t-d} - \xi_{t-d-1}) + W_{i,d}'(\eta_{t-d} - \eta_{t-d-1}) + \alpha^0_{t-d} - \alpha^0_{t-d-1} + (\alpha^1_{t-d} - \alpha^1_{t-d-1})\nu_i}_{\text{Effect of reduction in time at risk}} \quad + \\
&\underbrace{\theta^0_{d-1,t-d} - \theta^0_{d-1,t-d-1} + \nu_i\left[(\theta^1_{d-1,t-d} - \theta^1_{d-1,t-d-1})\right]}_{\text{Behavioral effects effects from a change in time at-risk}} \quad + \\
&\underbrace{(\theta^0_{d,t-d} - \theta^0_{d-1,t-d}) + \nu_i\left[(\theta^1_{d,t-d} - \theta^1_{d-1,t-d})\right]}_{\text{Behavioral effects holding fixed time at-risk}}
\end{aligned} \tag{M.13}$$

Appendix Figure M.3 shows the estimates of the behavioral effects of a marginal change in incarceration exposure. In this figure we use the simplified model specification that has a polynomial in D_i and an indicator for any incarceration sentence (instead of dummies for each month of exposure $\theta^0_{d,t-d}$ and $\theta^1_{d,t-d}$). Appendix Figure M.4 shows estimates of the same estimand using the more flexible specification of the control function. The additional precision provided by the polynomial specification is clearly demonstrated by comparing the estimates in Figure M.3 and Appendix Figure M.4. Although both show similar patterns, the estimates based on the polynomial specification are an order of magnitude more precise. The results using this decomposition are qualitatively similar to the previous decomposition.

Figure M.3: Decomposition based on decomposing the marginal increases in incarceration to incapacitation

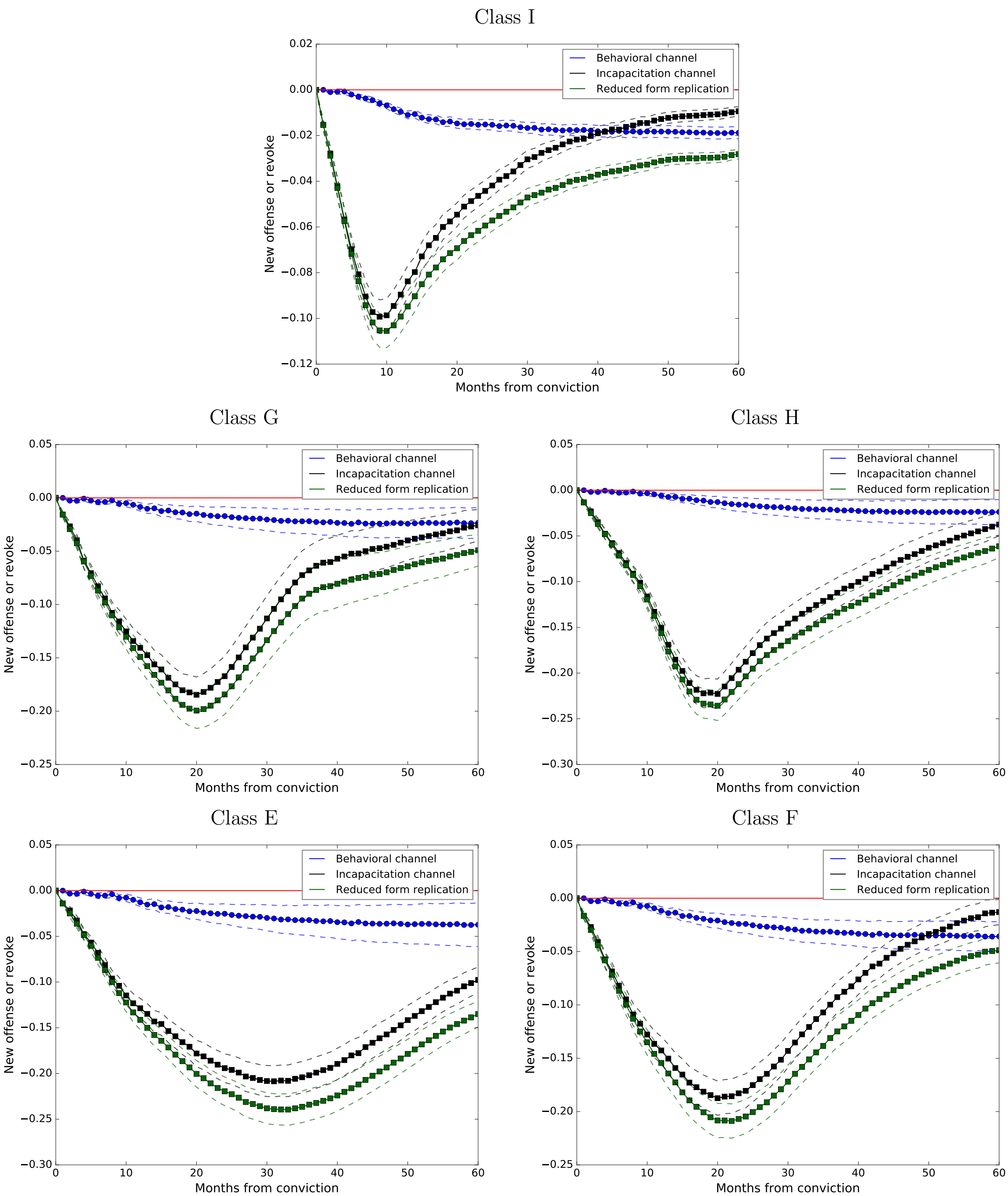
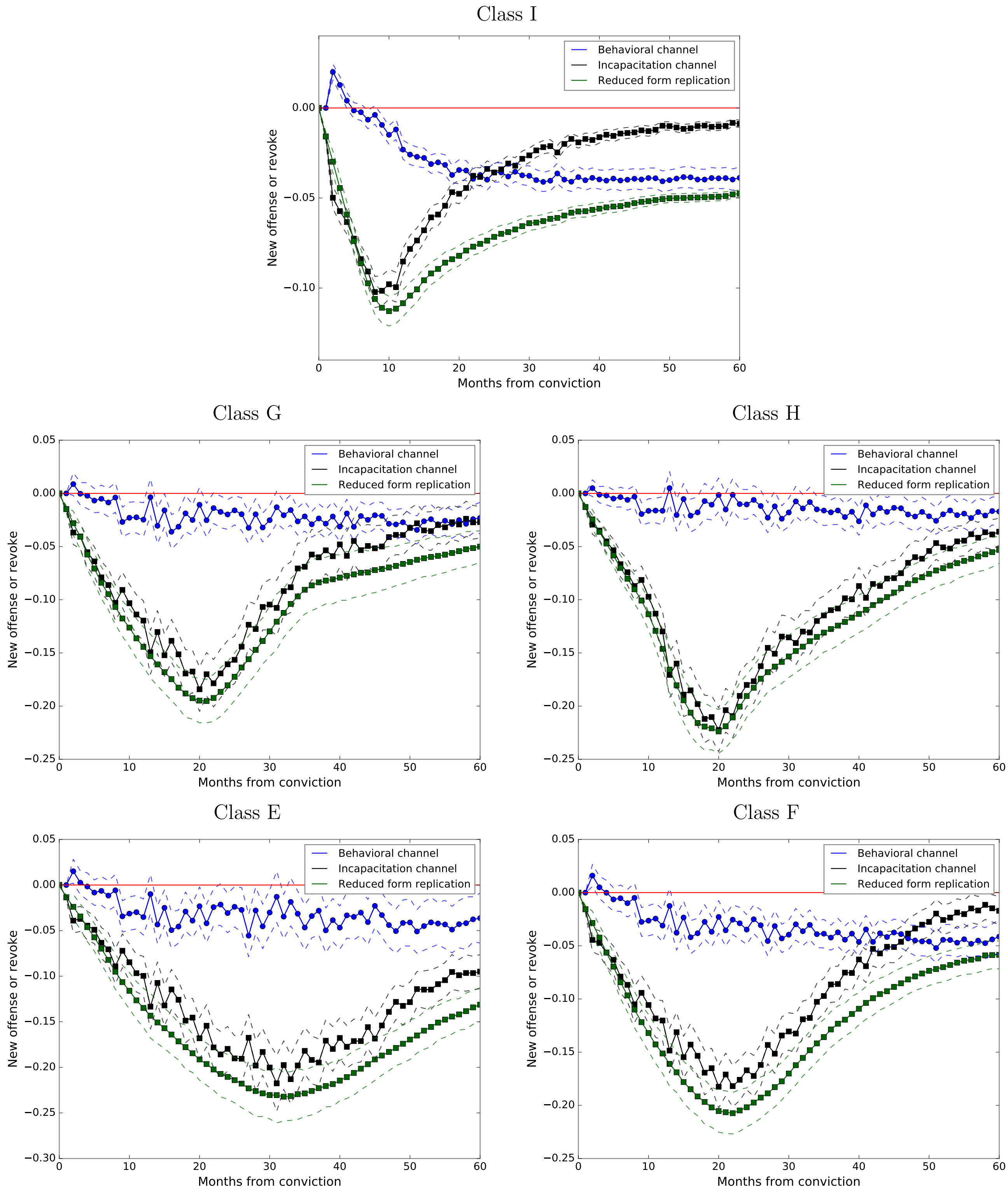


Figure M.4: HOLDING AT-RISK FIXED SHIFTING ONLY BEHAVIORAL - LEVELS



Notes: This figure shows the results of using the control function estimates to replicate and decompose the reduced form RD estimates of reoffending within t months from conviction The decomposition of the estimates to the incapacitation (black line) and behavioral (blue line) channels is done using the null of no behavioral effects. We first use the CF estimates to replicate the reduced form RD estimates (green line). Next we assume that there are no behavioral effects, i.e., we impose that the coefficients on all the incarceration variables/indicators are equal to zero, and replicate the RD estimates under this null (black line). The difference between the green and black lines is the unexplained part (blue line) in the estimates of the reduced forms and it can be attributed to the behavioral channel. We name this unexplained component the "behavioral residaul". We claculate SEs using a block bootstrap procedure with 500 iterations at the individual level to account for within-individual serial correlation.

M.5 Probation revocations as non-random censoring

Probation revocation poses a challenge for measuring reoffending that can lead to biased estimates of the impacts of incarceration. If mainly the non-incarcerated population are getting probation revocations and are being incapacitated in prison/jail without committing new offenses then incarceration estimates from at-risk will be upward biased, making incarceration look crime increasing even if it is actually not. In Section 2.3, we discussed how upper and lower bounds on the effects of incarceration on reoffending can be constructed; however, in practice this bounds are usually wide and are not informative enough. In this section, we present an enriched selection model with an additional selection corrections that captures, and corrects, for the non-random censoring of individuals due to technical probation revocations.

Consider the following triangular system of equations. The first point of selection is in which offenders are incarcerated and it is formulated according to Equation (7). The second point of selection is whether there is a non-random censoring of an individual, due to a technical probation revocation, before she commits a new offense.

This second selection point is described by the the following single index model:

$$P_{i,t}(d) = 1 \quad \text{if} \quad \mathbb{I}\left\{ \underbrace{\eta_{i,t-d}}_{\text{Unobserved heterogeneity}} \leq X_i' \delta_{t-d}^{l(i)} + \alpha_{t-d}^{l(i)} \nu_i + \gamma_{d,t-d} \right\} \tag{M.14}$$

where $R_{i,t}$ is an indicator whether individual i had a probation revocation prior to committing a new offense within t months of conviction; and $k = t - d$ is the number of months the offender was at-risk. The coefficient α_k^l can be interpreted as representing the correlation between ν_i and censoring due to a probation violation prior to committing a new offense, which is an unobserved relationship between selection into incarceration and the likelihood of getting a probation revocation.[44] Since ν_i is not observed it is integrated out

$$\Pr(R_{i,t} = 0 | X_i, Z_i, D_i) = \int_{a(Z_i^{l(i)}, D_i, X_i)}^{b(Z_i^{l(i)}, D_i, X_i)} \Pr(R_{i,t} = 0 | X_i, Z_i^{l(i)}, D_i, \nu_i = \nu) f_\nu(\nu) d\nu \tag{M.15}$$

where the upper ($b(Z_i^l)$ and lower ($a(Z_i^l)$ bounds of the integral are derived from the ordered choice model in Equation (7)

$$\begin{aligned} b(Z_i^{l(i)}, D_i, X_i) &= C_{D_i}^{l(i)}(Z_i^{l(i)}) - X_i' \gamma_0^{l(i)} \\ a(Z_i^{l(i)}, D_i, X_i) &= C_{D_i - 1}^{l(i)}(Z_i^{l(i)}) - X_i \gamma_0^{l(i)} \end{aligned} \tag{M.16}$$

The identification of $\eta_{i,k}$ relies on having a shifter (Z_i^l) that yields exogenous variation in $R_{i,t}$.

[44]Notice, that we assume that ν_i and $\eta_{i,k}$ are independent. The influence of ν_i on the likelihood that $R_i = 1$ is captured by directly including ν_i in the DGP of R_i.

The impacts of Z_i^l enter Equation (M.14) through the limits of the integral in Equation (M.15). Estimation is carried out using a simulated maximum likelihood procedure. To ensure that the instruments provide enough variation for estimating both selection corrections, we use the under-identification test proposed by Sanderson and Windmeijer (2016), which is a generalization of the test for under-identification that was proposed by Angrist and Pischke. When censoring is measured as a revocation prior to a new offense within three years of conviction, we reject the null of under-identification ($F = 78.77$ and $p \approx 0$ for censoring and $F = 219.9$ and $p \approx 0$ for duration of incarceration).

Next we describe a model for the relationship between incarceration and reoffending for non-censored offenders (i.e., conditional on $R_{i,t} = 0$). The conditional expectation of potential outcomes is

$$\mathbb{E}\left[Y_{i,t}(d)|X_i, Z_i^{l(i)}, \nu_i, R_{i,t}(d) = 0, \eta_{i,t-d}\right] = X_i'\xi_{t-d} + \theta^0_{d,t-d} + \theta^1_{d,t-d}\nu_i + \theta^2_{t-d}\eta_{i,t-d} \tag{M.17}$$

where $t - d$ is the number of months that offender i is at-risk to reoffend. The coefficients θ^2_{t-d} represent the unobserved correlation between the likelihood of reoffending and the probability of getting a technical probation revocation. If revocations are done at random then $\theta_{t-d} = 0$, if the offenders who are more (less) likely to commit crime are the ones who are censored then $\theta^2_{t-d} > 0$ ($\theta^2_{t-d} < 0$).[45]

By iterated expectations, Equation (10) yields that the conditional expectation of observed outcomes can be written as:

$$\begin{aligned}\mathbb{E}\left[Y_{i,t}|X_i, Z_i^{l(i)}, R_{i,t} = 0, D_i = d\right] &= X_i'\xi_{t-d} + \theta^0_{d,t-d} \\ &+ \underbrace{\theta^1_{d,t-d}\lambda^1_{t-d}\left(X_i, Z_i^{l(i)}, d\right)}_{\text{Correction for selection into incarceration}} + \underbrace{\theta^2_{d,t-d}\lambda^2_{i,t-d}\left(X_i, Z_i^{l(i)}, 0, d\right)}_{\text{Correction for non-censored offenders}}\end{aligned} \tag{M.18}$$

where the selection correction $\lambda^2_{t-d}\left(X_i, Z_i^{l(i)}, 0, d\right)$ is a variant of the inverse Mills ratio

$$\begin{aligned}\lambda^2_{t-d}\left(X_i, Z_i^{l(i)}, 0, d\right) &\equiv \int_{a(Z_i^{l(i)}, D_i, X_i)}^{b(Z_i^{l(i)}, D_i, X_i)} \mathbb{E}\left[\eta_{i,t-d}|R_{i,t} = 0, X_i, Z_i^{l(i)}, D_i = d, \nu_i = \nu\right] f_\nu(\nu)d\nu \\ &= \int_{a(Z_i^{l(i)}, D_i, X_i)}^{b(Z_i^{l(i)}, D_i, X_i)} \mathbb{E}\left[\eta_{i,t-d}|\eta_{i,t-d} \leq X_i'\delta^{l(i)}_{t-d} + \alpha^{l(i)}_{t-d}\nu + \gamma_{d,t-d}, X_i, Z_i^{l(i)}, D_i = d\right] f_\nu(\nu)d\nu \\ &= \int_{a(Z_i^{l(i)}, D_i, X_i)}^{b(Z_i^{l(i)}, D_i, X_i)} \frac{\phi\left(X_i'\delta^{l(i)}_{t-d} + \alpha^{l(i)}_{t-d}\nu + \gamma_{d,t-d}\right)}{1 - \Phi\left(X_i'\delta^{l(i)}_{t-d} + \alpha^{l(i)}_{t-d}\nu + \gamma_{d,t-d}\right)} f_\nu(\nu)d\nu\end{aligned} \tag{M.19}$$

[45]The new selection correction reveals what type of bias is introduced to the incarceration estimates by conditioning on offenders who did not have their probation revoked prior to committing a new offense.

Also notice that the generalized residuals from the probit model in Equation (M.14) are

$$
\begin{aligned}
&R_{i,t} \cdot \int_{a(Z_i^{l(i)},D_i,X_i)}^{b(Z_i^{l(i)},D_i,X_i)} \frac{\phi\left(X_i'\delta_{t-d}^{l(i)} + \alpha_{t-d}^{l(i)}\nu + \gamma_{d,t-d}\right)}{\Phi\left(X_i'\delta_{t-d}^{l(i)} + \alpha_{t-d}^{l(i)}\nu + \gamma_{d,t-d}\right)} f_\nu(\nu)d\nu \qquad \text{(M.20)}\\
&\quad - (1 - R_{i,t}) \cdot \int_{a(Z_i^{l(i)},D_i,X_i)}^{b(Z_i^{l(i)},D_i,X_i)} \frac{\phi\left(X_i'\delta_{t-d}^{l(i)} + \alpha_{t-d}^{l(i)}\nu + \gamma_{d,t-d}\right)}{1 - \Phi\left(X_i'\delta_{t-d}^{l(i)} + \alpha_{t-d}^{l(i)}\nu + \gamma_{d,t-d}\right)} f_\nu(\nu)d\nu
\end{aligned}
$$

M.5.1 Identification

Next we describe the variation that is necessary for identification when the control function approach includes a second selection correction for technical probation revocations that occur prior to committing a new offense.

The model in Equation (M.18) is also over-identified, but now we require variation from more than one instrument since there are two selection correction in the model. Similar to model (11) the identification of $\theta_{d,k}^1$ and $\theta_{d,k}^2$ relies on variation in Z_i^l given $X_i = x$, $D_i = d$, $k = t - d$ and $R_{i,t} = 0$

$$
\begin{aligned}
&\mathbb{E}\left[Y_{i,t}|X_i = x, D_i = d, k = t - d, Z_i^l = 1\right] - \mathbb{E}\left[Y_{i,t}|X_i = x, D_i = d, k = t - d, Z_i^l = 0\right] \qquad \text{(M.21)}\\
&\quad = \theta_{d,k}^1\left(\lambda^1(x,1,d) - \lambda^1(x,0,d)\right) + \theta_{d,k}^2\left(\lambda^2(x,1,0,d) - \lambda^2(x,0,0,d)\right)
\end{aligned}
$$

the above derivation can be done for each of the five instruments yielding an over-identified model of a system of five equations with two unknowns.