Title
Essays in Labor Economics and the Criminal Justice System

Permalink
https://escholarship.org/uc/item/3sd0f741

Author
Shem-Tov, Yotam

Publication Date
2019

Peer reviewed|Thesis/dissertation

Essays in Labor Economics and the Criminal Justice System

by

Yotam Shem-Tov

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Patrick Kline, Chair
Professor David Card
Professor Steven Raphael
Professor Christopher Walters

Spring 2019

Essays in Labor Economics and the Criminal Justice System

Abstract

Essays in Labor Economics and the Criminal Justice System

by

Yotam Shem-Tov

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Patrick Kline, Chair

This dissertation investigates key aspects of the U.S. criminal justice system.

The first chapter studies different methods of providing a legal counsel to low-income criminal defendants. Most criminal defendants in the U.S. cannot afford to hire an attorney. To provide constitutionally mandated legal services, states commonly use either private court-appointed attorneys or a public defender organization. This paper investigates the relative efficacy of these two modes of indigent defense by comparing outcomes of co-defendants assigned to different types of attorneys within the same case. Using data from San Francisco, I show that in multiple defendant cases public defender assignment is plausibly as good as random. I find that defendants who have been assigned a public defenders obtain more favorable sentencing outcomes.

The second chapter investigate the causal effect of incarceration on reoffending using discontinuities in state 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.

The third chapter parses the non-linear and heterogeneous effects of incarceration on post-release criminal behavior, I develop an econometric model of sentencing length and recidivism. This model model allows for Roy-style selection into sentencing on the basis of latent criminality. I 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 that I study. The 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.

To my mother.

Contents

List of Figures

List of Tables

List of Tables

Acknowledgments

I am grateful to my main advisor, Patrick Kline, for all his guidance, help, and constant support throughout the years at Berkeley. Pat has been a role model both as an academic and as an advisor with endless patience and dedication to his students. I am also grateful to David Card, Jasjeet Sekhon and Christopher Walters for their encouragement at all stages of my PhD. Steven Raphael, Justin McCrary and Emmanuel Saez have provided invaluable feedback to improve my work. I am deeply lucky to have had the chance to get such training and guidance. The Berkeley faculty taught me that in a seminar the aim is to give constructive feedback to the presenter, to be patient, professional, and most importantly always be generous with your time. David is tremendously busy and yet always makes time to advise students, not because he needs to but because he wants to. I will always be indebted to him for all his honest, direct and extremely helpful feedback. I also have learned a lot and benefited greatly from conversations with Alan Auerbach, Hilary Hoynes, Enrico Moretti, Jesse Rothstein, Fred Finan, and Danny Yagan. I am grateful to my co-author and collaborator, Evan K. Rose, for his support and help throughout the PhD, and to Vicky Lee and Camille Fernandez for all of their tremendous help.

My time in Berkeley has been wonderful thanks to many friends and colleagues, including Alessandra Fenizia, Jonathan Schellenberg, Avner Shlain, Daniel Haanwinckel, Itai Trilnick, Zarek Brot-Goldberg, Maxim Massenkoff, Jonathan Holmes, Francis Wong, Zachary Bleemer and John Loeser. I especially thank Nicholas Li for endless discussions and econometrics talks throughout our years in Berkeley.

My mother, Debbie Bernstein, has supported me throughout my life. Her faith and support have been essential for my success as a person and an academic. You are my inspiration and a role model, as a person and as an academic scholar.

Lastly, I would like to thank Juliana, my wife and best friend, for all her endless help, discussions, and fruitful exchange of ideas and thoughts. You taught me so many things during the last six years and all in endless patience and love. Your love and selfless support have made this journey the wonderful experience that it was.

Chapter 1

Make-or-Buy? The Provision of Indigent Defense Services in the U.S.

I Introduction

Low-income individuals facing criminal charges in the U.S. have a constitutionally protected right to legal counsel from an attorney who is appointed and compensated by the state. Legal counsel is essential since "the average defendant does not have the professional legal skill to protect himself when brought before a tribunal with power to take his life or liberty" (Johnson v. Zerbst, 1938). Among felony defendants, 80% require the assistance of such services (Harlow, 2001). While empirical research has focused on the role of judges in determining case outcomes (Anderson, Kling, and Stith, 1999; Mustard, 2001; Abrams, Bertrand, and Mullainathan, 2012; Yang, 2015; Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan, 2017; Arnold, Dobbie, and Yang, 2018; Cohen and Yang, 2018), the importance of defense attorneys is relatively underexplored. Differences across attorneys in defendants' case outcomes can have long-lasting impacts. A felony conviction or incarceration sentence can harm earnings, employment, and educational attainment (Grogger, 1995; Raphael, 2011; Aizer and Doyle, 2015b; Mueller-Smith, 2015; Agan and Starr, 2018).

The main challenge in evaluating the performance of PD relative to CA is that the usual mechanism of assigning an indigent (i.e., low-income) defendant to a PD is *not* random and can vary across jurisdictions. While defendants cannot manipulate the process, the judge, court, and public defender's office can potentially influence the assignment procedure. Indeed, I find that defendants represented by a PD are substantially different in their observable characteristics than those represented by a CA.

To overcome this challenge, I use detailed court records from San Francisco and employ a new identification strategy of comparing co-defendants within the same case. In multiple defendant cases the PD office does not represent co-defendants to avoid inherent conflicts of interest (Allison, 1976; Lowenthal, 1978; Moore, 1984). In general, the within-case assignment of defendants to a PD does not have to be random. However, I show that in San

Francisco, the decision of who will be assigned a PD in multiple defendant cases is as good as random. The within-case assignment to a PD is not correlated with defendant characteristics such as race, age, criminal history, and charge severity. Selection on unobserved factors is possible although unlikely, since these omitted variables need to be correlated with both case outcomes and PD assignment, but uncorrelated with criminal history, charge severity, age, and race. I exploit this natural experiment to quantify the causal effect of being assigned a PD relative to a CA on case outcomes.

I find that co-defendants assigned to a PD generally obtain more favorable sentencing outcomes. Co-defendants assigned to a PD have a lower probability of both conviction (6.4%), and prison sentence (22%), as well as a shorter expected imprisonment term (10.8%). There is no evidence of heterogeneity with respect to the defendant's demographic characteristics (e.g., gender, race); however, there is substantial heterogeneity with respect to criminal history and charge severity. Defendants who face more severe criminal charges (e.g., felony vs. misdemeanor) and have a longer criminal history are the ones driving the results. This implies that as the likelihood of incarceration is higher, due to having a longer criminal history or facing more severe charges, the effect of having a better legal counsel will be larger.

I also find evidence that cross case comparisons can potentially yield biased estimates that are confounded by other factors. For example, in San Francisco, the effect of being assigned a PD on incarceration length changes from -35.7% to -18% when comparing across cases and including single defendant cases. This is evidence that individuals who are likely—based on observables—to receive a shorter incarceration sentence are the ones assigned to a PD. Unlike the *across* cases comparison, in the within-case comparison the estimates with and without controls are similar, -10.8% and -10.9%; and are *smaller* in magnitude. This implies that studies that simply compare these two attorney types will tend to overestimate the efficacy of PDs relative to CAs. Focusing on multiple defendant cases raises a concern about the external validity of the results to other contexts. Although the vast majority of cases involve only a single defendant, one of the most frequent scenarios in which a conflict of interest arises is in multiple defendant cases. The results of the within-case comparison are therefore of direct policy relevance for every jurisdiction that has a PD office.

To understand the external validity of these results from San Francisco to other jurisdictions, I conduct a similar analysis using data from federal district courts. In contrast to San Francisco, in federal courts, I document both across and within-case selection. Within a multiple defendant case, the order in which defendants are listed on the indictment is correlated with both PD assignment and the defendant's culpability. To overcome this issue, I condition on the defendant's order of appearance on the indictment using an order-specific fixed effect. Once the defendant's position on the indictment (e.g., first, third) is taken into account, the assignment to a PD or a CA can be treated as if it were done independently of the defendant's culpability.

The findings are qualitatively similar to San Francisco, although the magnitude of the effects is much smaller. While the probability of being convicted is not affected by the assignment to a PD, the expected prison sentence is 4.64% shorter and the probability of any prison sentence is 1% lower. This is partly explained by the fact that the overall rates of conviction and incarceration are dramatically high in federal courts: 95% of defendants are convicted and 86% are sentenced to prison relative to 59.8% and 6.3% in San Francisco.[1]

Having established that PDs provide better legal representation, I turn to investigate possible mechanisms. On explanation is that individuals who sort to work as PDs are different than those who elect to serve as CAs. Another channel can be the organizational norms, mentoring, and other resources that are more available to PDs. I document large differences in the observable characteristics between the two attorney types: PDs are younger, demographically more diverse (higher share of females and non-whites), graduate from B.A. and J.D. programs in higher-ranked institutions, and have more court experience. These differences provide descriptive evidence about the adverse selection into the pool of attorneys who choose to accept indigent defense appointments relative to the attorneys who select to work for a PD office. One policy implication can be to established an alternate PD organization for situations in which the main PD cannot represent an individual due to conflicts of interest as is the case in Los Angeles.

This paper contributes to the nascent literature on the importance of the defense attorney. Previous studies utilizes various empirical methods to evaluate the importance of the attorney's characteristics, type, and compensation scheme on defendants' case outcomes (Abrams and Yoon, 2007; Anderson and Heaton, 2012; Agan, Freedman, and Owens, 2018). The study also contributes to a large body of literature on whether the state should "make-or-buy" public services. Other examples of such decisions range from schools (Abdulkadiroglu, Pathak, and Walters, 2015) to police (Cheng and Long, 2017) and prison (Mukherjee, 2017). Weak populations (e.g., prisoners, criminal defendants) are especially vulnerable to the privatization of public services (Hart, Shleifer, and Vishny, 1997).

Most related to this study is Anderson and Heaton (2012) who exploit the initial random assignment of defendants in homicide cases in Philadelphia to CA and PD to compare between the two. They find that being assigned a PD reduces the defendant's sentenced imprisonment time by 31% but has *no* effect on the probability of conviction. This paper extends Anderson and Heaton (2012) in several directions.

The first is external validity. I evaluate PDs and CAs in a range of different offenses and not only in homicide trials, which are a rare and unrepresentative procedure, constituting only 0.1% of arrests in the U.S. in 2016. Unlike Anderson and Heaton's results, I find that assignment to a PD causes a significant reduction in the probability of conviction. I also find that the majority of the differences in prison sentences between attorney types is driven by felony cases and defendants with a prior criminal history who face a higher probability of incarceration. Since homicide is the most severe offense, it is expected that the attorney type effects will be the largest in these cases, which also explains why my estimated effects on sentencing length are lower (a 10.8% relative to a 31% reduction). Second, I document

[1] In San Francisco, the share of defendants sentenced to prison is 6.3% and the share who are sentenced to either jail or prison is 16%. These rates of conviction and incarceration are similar to the ones in other jurisdictions. For examples, see Table 2 in Feigenberg and Miller (2018).

how PDs are different from CAs in their observable characteristics and quantify how much of the estimated effects can be accounted for by these attorney-observable characteristics. The differences in attorney characteristics emphasizes that selection of attorneys into PD organizations relative to accepting cases as a CA can account for a meaningful share of the differences in efficacy.[2]

Abrams and Yoon (2007) and Agan et al. (2018) do not compare PD and CA, but rather investigate how the characteristics and compensation scheme of the defense attorney can impact the defendant's case outcomes. Abrams and Yoon (2007) utilizes the quasi-random assignment of attorneys to cases within the public defender's office in Nevada, to estimate the impact of attorney characteristics such as experience, gender, and race on case outcomes. They find that these characteristics are important in predicting the defendant's case outcomes. Agan et al. (2018) compares the case outcomes of individuals represented by a privately retained counsel relative to those represented by a private court appointed attorney (i.e., CA). They find that wage incentives are an important factor in explaining defendants' case outcomes.

My findings regarding differences in attorney quality inform the policy discussion on the high rates of incarceration in the U.S. For instance, can we reduce the share of the population that is imprisoned by providing better legal assistance? My estimates indicate that assignment to a PD decreases the likelihood of being sentenced to prison by 22% in state courts, but has a negligible 1% impact in federal courts. This suggests that, at least in state courts, changing the method of provision can have a lasting impact on the share of defendants who are incarcerated as well as on reoffending patterns and thus long-term relationships with the criminal justice system. However, better legal representation can also cause an increase in crime if more lenient case outcomes enable defendants to reoffend easier or sooner (Ater, Givati, and Rigbi, 2017).

The remainder of the paper is organized as follows. Section II describes the data. Section III presents the identification strategy and verifies the validity of the multiple defendant design. Section IV presents the empirical findings. Section VI, zooms out of San Francisco and provides an analysis on the importance of the attorney type (e.g., PD, CA) in federal courts. Section VII concludes and briefly suggests avenues for future research.

[2] Other studies include Iyengar (2007) and Roach (2014) both argue that by using a data-driven procedure it is possible to detect location-year pairs in which the attorney type assignment was done at random. In Appendix A I describe the implementation of this data-driven procedure in federal district courts and show evidence that it does not succeeds in detecting jurisdictions that randomly assign defendants across attorney types. My estimates of assignment to a PD in this paper are substantially different in magnitude than the ones reported by both Iyengar (2007) and Roach (2014).

II Data

Data sources and sample construction

I use administrative records from the court system in San Francisco for all cases terminated between February 2006 and March 2016. The data contains sentencing outcomes such as whether the defendant was convicted, and if so the length of prison sentence and length of probation, as well as a detailed description of the filed charges ranging from broad characteristics such as a felony or a misdemeanor to more granular information on the specific statute and title of the offense. I also calculated SC and BCS codes for each charge, which are classifications of offenses to broad categories of severity that are commonly used by the California Department of Justice.[3] Basic demographic information on the defendant such as race, sex, and age is available, and I use names to infer Hispanic origin using data from the 2000 Census.

Defendants often face multiple charges at the time of disposition for offenses that took place at different times. For example, an individual can be charged with offense A and then be released on bail; while awaiting trial he can then be additionally charged with offense B. The disposition of both charges can take place at the same time. In the above scenario, if the defendant is indigent, the attorney who represented him for charge A will also be his counsel for charge B. Therefore, attorney assignments are based on the initial charging of each case. I group charges together into cases based on whether the conviction, offense, or charging date fall within a certain time window (e.g., 20 days) of each other for a given individual. I then define the initial attorney type assignment as the first attorney that represented the defendant within a case.[4]

Two individuals are defined as co-defendants if they have the same police incident number, i.e., if they have been arrested for the same underlying criminal event. For the rest of the paper I refer to individuals with the same police incident number as co-defendants in the same case. In sections II and III, I discuss the distribution of defendants across attorney types in multiple and single defendant cases (i.e., police incidents). For the main analysis, I restrict attention to criminal cases in which the defendant was initially represented by an appointed counsel: either a PD or a CA.

Defendants and assignment across attorney types in San Francisco

The provision of indigent defense services varies widely across jurisdictions in the U.S. This includes both the type of attorney (CA or PD) as well as the level of compensation the attorney receives. In San Francisco, the public defender's office was established in 1921 and represents the majority of indigent defendants. The CA attorneys in San Francisco (known as

[3]See `https://oag.ca.gov/law/code-tables`

[4]The choice of using no time window (i.e., grouping together charges with exactly the same conviction, offense, or charging date) or a window of 5, 10 or 20 days has no impact on the results. The effects are similar regardless of this choice.

"conflict attorneys"), are considered to be professionals who provide competent legal counsel to their clients. They are not obliged to represent clients and the court compensates them for their work. CA attorneys must satisfy strict requirements to be eligible to receive indigent defense appointments from the court.

Indigent defendants are generally assigned to the PD office in San Francisco unless there is a conflict of interest; only then are cases assigned to CAs. For example, if the PD office previous represented a witness in the case.[5] Figure 1.1 shows the distribution of defendants across attorney types in San Francisco from 2006 to 2015. In single defendant cases (Panel A), the vast majority of indigent defense representation is done by the PD office; however, within multiple defendant cases the division is almost equal (Panel B). This prevalence of CAs in multiple defendant cases results from the fact that the PD office in San Francisco avoids representing more than one defendant within a case as is discussed in more detail in Section III.

Table 1.1 presents summary statistics on criminal defendants in San Francisco. Column (2) includes all cases with more than one defendant and column (3) all cases with at lease one defendant that is represented by a PD and another by a CA such that both a PD and a CA are present in the case. Approximately 50% of the defendants are Caucasian and African-Americans are overly represented.[6] The share of African-Americans and females is higher in multiple defendant cases (columns (2) and (3)) relative to single defendant cases (column (1)). The average age in multiple defendant cases is lower than in single defendant cases 32 relative to 35. Multiple and single defendant cases vary also in the severity of the charges: 82.9% include a felony charge relative to 51.8% respectively; and the probability to be incarcerated in prison (jail) is higher (lower) in multiple relative to single defendant cases. In almost a quarter of the cases the charges are eventually dropped. Multiple defendant cases that include both a PD and a CA (column 3) are the majority of multiple defendant cases and are similar to the overall sample of multiple defendant cases (column 2) in defendant demographics, charge severity measures and case outcomes.

III Identification strategy: Conflict-of-interest considerations in cases of multiple defendants

In multiple defendant cases, the public defender's office is usually constrained to represent only a single defendant due to potential conflicts of interest (Moore, 1984; Allison, 1976;

[5]Conflicts of interest can occur under various circumstances. Other examples include, multiple defendant cases in which the interests of the different individuals can contradict each other and they will each require a separate legal counsel. Another example is the attack that took place in Charlottesville, Virginia (link to article below). The PD office could not represent the accused in the attack since certain members of the office had family members who were wounded in the assault, and the defendant was assigned a CA. `https://www.theguardian.com/us-news/2017/aug/14/james-fields-charlottesville-driver-murder-charge?CMP=Share_AndroidApp_Gmail`.

[6]The share of African-Americans in the population of San Francisco was approximately 6% in 2010.

Lowenthal, 1978; Prado, Altman, Aprile, Clark, Davis, Dennis, Everett, Hillier, Revercomb, and Ogletrel, 1993). The Committee to Review the Criminal Justice Act, 1991–1992, determined that a "defender organization cannot properly undertake the representation of more than one defendant in a multi-defendant prosecution because a conflict of interest almost invariably results." The review committee specifically states that "private attorneys provide representation in multi-defendant and other cases in which representation by the federal defender could potentially create a conflict of interest."

In such circumstances, usually a PD is assigned to one of the indigent defendants, and the others are appointed to CAs. Figure 1.2 shows how conflict of interest considerations impact the attorney type assignment in multiple defendant cases. Panel A shows the average number of defendants who are represented by each type of attorney (e.g., PD, CA) by the number of defendants in the case and Panel B shows the share of defendants within a case that are assigned to each type of attorney. The figure clearly validates the conflict-of-interest hypothesis that the PD organization will usually not represent more than one defendant within a multiple defendant case. In federal courts similar patterns emerge (Appendix Figure A.4) as is discussed in Section VI.

The within-case comparison can be viewed as matching together similar units and then "randomly" flipping a coin to assign some to the treatment and the others to the control. Naturally, the setting, rather than statistical methods designed to optimize covariate balance, creates the matches. The matches are pre-determined outside of the control of the researcher. Thus, covariate balance can be used as a testable implication to validate the assumption that treatment was exogenously assigned within each case.[7]

Overcoming selection in the assignment of defendants between PDs and CAs

I begin by documenting extensive selection in the assignment of defendants between PDs and CAs, which is essential to overcome in order to understand whether PDs and CAs provide the same level of legal representation. If the cases that are assigned a PD are different in their severity and complexity as compared to those that are assigned a CA, then these differences need to be taken into account when the case outcomes are compared. To summarize the differences in the charges that defendants who are assigned a PD relative to a CA are facing, I use covariate indices that are based on a Oaxaca decomposition. In Appendix A, I describe

[7]As the assignment to a PD or a CA is usually not done by a flip of a coin, there is a concern that in cases involving numerous defendants (e.g., 30, 50, 100) the independence assumptions may be less plausible. Many defendants' cases may begin to inherit selection problems characteristic of the full sample since there may be differences in defendants' probability of assignment to a PD. To mitigate concerns of selection bias in multiple defendant cases, I limit the cases in our sample to have no more than 10 defendants. Relaxing this assumption to cases involving no more than 5 or 20 does not change the results of the paper. The constraint is binding only in federal courts. In San Francisco, the vast majority of multiple cases are of co-defendants, two defendants within a case, and there is a small number of cases with more than four defendants. There are no cases with more than 10 defendants in San Francisco.

the exact construction of the covariate indices that are used both to document selection and to test for balance within a multiple defendant case. In both San Francisco and federal district courts, I observe offense codes that are highly predictive of the case outcomes but are too numerous to show comparisons for each category separately. The dimensional reduction that is conducted using the summary covariate indices allows me to present one summary measure that includes imbalances in demographics, charge severity measures and criminal history all at once

$$X_i = \beta \cdot PD_i + \alpha_{j(i)} + e_i, \tag{1.1}$$

where the β coefficient is the average difference in characteristic X_i across defendants represented by a PD relative to a CA. When case fixed effects $\alpha_{j(i)}$ are not included, the β coefficient is exactly the difference in means, and when they are included it is the within a case difference in means. A cross-case comparison, when fixed effects are not included, can be sensitive to omitted-variable bias if there is selection in the type of cases that are assigned a PD relative to a CA.

Table 1.2, columns (1) and (2), shows clear evidence of selection in the assignment of defendants between PD and CA. The CA attorneys represent significantly more African-Americans, females, and defendants who are facing more severe offenses and face a longer expected imprisonment time if convicted. This pattern of non-random sorting is the result of two factors. The first is that in San Francisco, the PD office handles the vast majority of cases, which are mostly not felony cases. Second, in cases with more severe charges there is a higher likelihood of a conflict of interests (e.g., between co-defendants), which leads to a higher proportion of defendants who are assigned to a CA among defendants facing felony-level charges.

In column (3), I restrict attention to multiple defendant cases with both PD and CA; however, I do not take into account variation in case-level characteristics (e.g., number of defendants). The differences between attorney types in columns (2) and (3) are based on a comparison of defendants across court-cases. In multiple defendant cases, a cross case comparison can provide a false impression as the number of CAs changes with the number of defendants in the case while the number of PDs is approximately fixed at one. When the severity of the charges increases with the number of defendants, it is necessary to adjust for case fixed effects, i.e., conduct a within-case comparison, to obtain a reliable estimate of the differences in charge characteristics between defendants who are assigned a PD relative to a CA *within* a case.

Table 1.2, columns (4) and (6), show that *within* a multiple defendant case the treated and control units are comparable in demographic characteristics and charge severity measures. The adjusted difference in means using case level fixed effects, columns (4) and (6), shows that the differences between treated and control units are not statistically significant and are especially small relative to the baseline means of each measure (described in Table 1.1).

Figure 1.3 provides a visualized summary of the estimates in Table 1.2. Each point on the figure is a t-statistic of the β coefficient in equation (1.1). The figure visualizes clearly how the selection in the attorney type assignment goes away once the comparison is conducted within a case.

Furthermore Appendix Figure A.1 reports the results of a joint F-test for whether the controls are predictive of the attorney type assignment. The figure reports the observed value of the test statistics, the F-stat, and its likelihood under the null distribution of random assignment of defendants to attorney types. The null distribution was generated by a Monte-Carlo simulation with 1,000 random permutations of the PD assignment within a case, in multiple defendant cases, and across cases in single defendant cases. It is clear that in single defendant cases the assignment is not done at random; however, in multiple defendant cases there is no evidence of sorting within a case and we cannot reject the null hypothesis of random assignment.

Finally, Appendix Table A.1 documents substantial within-case variation in observables. For example, in 33.8% of the cases there is at least one defendant with a prior arrest and one without. In 22.2% of the cases there is at least one black and one non-black defendant. The above analysis shows that this variation is *not* correlated with the within-case attorney type assignment, which supports the assumption that the multiple defendant scenario can be considered a natural experiment with exogenous PD assignment within a case.

IV The attorney type effect on case outcomes

The main objective of this paper is to estimate the causal effect of assignment to a PD relative to a CA on the defendant's case outcomes. I argue that by conditioning on a sample of multiple defendant cases with both a PD and a CA within a case, the within-case assignment to a PD can plausibly be considered as good as random. As outlined in further detail in Section III, a PD office is constrained to representing only one client in a multiple defendant case due to potentially conflicting interests between co-defendants. In Section III, the balance tests show that within a case, the defendants with a PD and those with a CA are not observably different. Therefore comparing outcomes using *within-case* variation can limit selection biases that may arise from comparing outcomes *between cases*.

Let $\mathrm{PD}_i \in \{0,1\}$ be an indicator of whether defendant i was first assigned a PD, and let Y_i denote some sentencing outcome of interest (e.g., length of imprisonment). A standard causal model that relates the defendant's attorney type to his case outcomes is:

$$Y_i = \beta \cdot \mathrm{PD}_i + X_i'\Gamma + \alpha_{j(i)} + \epsilon_i, \tag{1.2}$$

where $j(i)$ is a mapping from defendant i to court case number j, X_i is a vector of observable pre-treatment variables that include measures of the severity of the filed charges (e.g., offense codes) and the type of charges (e.g., felony, misdemeanor), the demographic characteristics of the defendants and their criminal history; and β is the effect of assignment to a PD on case outcome Y_i.[8]

[8] The β coefficient should be interpreted as the effect of being assigned a PD relative to a CA *given* that

Table 1.3 reports the estimation results. In the full sample with both single and multiple defendant cases, individuals who are first assigned a PD are sentenced to a shorter prison term by nearly 33.1% relative to those assigned a CA. This unadjusted difference falls to 18% with the inclusion of controls, which suggests that a naïve comparison can be influenced by selection bias in the assignment of defendants to different attorney types. Differences in *observable* defendant and charge characteristics explain a substantial share of the sentencing differences between those who are assigned a PD vs. a CA. Altonji, Elder, and Taber (2005) show that differences between the covariate-adjusted and unadjusted estimates can be interpreted as a measure of selection due to omitted variables, which re-enforces the claim that a simple regression that relies upon a strong, unverifiable conditional independence assumption will not identify a causal relationship.

Table 1.3, columns (4)–(6), shows that within multiple defendant cases, those assigned a PD are sentenced to a 10.5% *shorter* prison term relative to their co-defendants who are represented by a CA. The estimate with covariate adjustment (a 10.7% shorter prison term) is not statistically different from the unadjusted estimate, which stands in contrast to the differences in estimates with and without covariate adjustments in the full sample that includes single defendant cases. Column (6) reports the results (10.1%) once controlling for prior representation by a PD, which also does not impact the estimate. Assignment to a PD also decreases the probability of conviction by 6.4% (3.9pp) and any prison time by 22% (1.8pp) relative to the mean rate of imprisonment. I find the attorney type of the defendant has no statistically significant effect on the sentenced jail term (-0.3pp) or the probability of being released on bail (-0.018pp).

Figure 1.4 summarizes the magnitude of the estimated attorney type effects within multiple defendant cases relative to the baseline mean of each sentencing outcome. The right plot presents confidence intervals for the estimated effects. The left plot illustrates the likelihood of the observed estimated effects relative to a null distribution in which the attorney type assignment has no effect. The null distribution was generated by a Monte-Carlo simulation with 1,000 random permutations of the PD assignment within a case. The black dots indicate the values of the coefficient that were obtained by a random permutation and darker areas represent values of the coefficient that are likely to be observed under random chance when the attorney type has no effect. The red triangles indicate the observed values of β in the data. The results suggest that overall PDs obtain significantly more favorable case outcomes for their clients in a range of sentencing outcomes. The largest effects are on the defendant's prison term, the probability of being sent to state prison at all, and the probability of conviction. The estimated effect on the probability of being sent for any period of time to a state prison is a 22% decrease relative to the baseline mean.

these are the two options of representation. I am *not* comparing the effect of being assigned a CA relative to a counterfacual that both defendants are represented by the same PD office. It is important to note, that an alternative to CAs is to have multiple PD organizations that handle situations at which there are conflicts of interest. For example, this is the case in Los Angeles, where there is an alternate PD office that handles situations at which the PD office cannot represent an indigent defendant such as conflicts of interest (`http://apd.lacounty.gov/`).

Next I investigate whether the attorney type assignment has heterogeneous effects across defendants, by interacting PD_i with various defendant characteristics:

$$Y_i = \beta \cdot \text{PD}_i + \gamma \cdot \text{PD}_i \times C_i + X_i'\Gamma + \alpha_{j(i)} + \epsilon_i, \tag{1.3}$$

where C_i is one of the elements of X_i'.[9]

Figure 1.5 reports the results of the heterogeneity analysis for three different outcomes. The figure shows that defendants who are facing more severe charges (felony vs. misdemeanor) and those who have a longer criminal history are the ones who stand to benefit the most from being represented by a PD relative to a CA. The results in Figure 1.5 illustrates that the interactions of defendant charge severity and criminal history with attorney type are unlikely under random chance and are statistically significant when permutation inference is used. The heterogeneous effects are studied with respect to the probability of being imprisoned, the length of imprisonment, and the probability of conviction. Appendix Table A.2 reports the P-values of the observed estimates of γ relative to their distribution under random chance, which is visualized in Figure 1.5.

In my empirical setting, the use of permutation inferences seems to provide accuracy gains when evaluating the likelihood of observing the estimated effects (i.e., the γ coefficients) relative to the null hypothesis that the attorney type assignment has no impact on case outcomes. In Appendix A, I compare permutation inference and the usual cluster-robust standard errors in terms of power in my context and find that permutation inference can have substantially higher power to reject the null when it is false.[10]

Finally, providing defendants with a higher quality of legal representation can lead to fewer defendants being sent to prison, which might cause an increase in crime as they will not be incapacitated (Ater et al., 2017). I estimate equation 1.2, where the outcome, Y_i, is recidivism within a certain period of time (e.g., 10 weeks) from the date of disposition. Figure A.2 reports the β estimate, which is not statistically significant, but is positive and increases until 60 weeks, at which point the coefficient stabilizes, and at around 120 weeks it starts to decline toward zero.[11]

These results are policy relevant since from a constitutional perspective, if defendants in the same case have been assigned attorneys with varying levels of legal expertise to such a degree that it influenced their sentencing outcomes, then there is a concern about a violation of the defendants' Sixth Amendment rights.

[9] The above specification includes all the main effects of the interaction term.

[10] This is *not* a general result that can be applied to other contexts, but rather specific to my application. In Appendix A, I use Monte-Carlo simulations to compare the power of the two methods of conducting inference. All the details are described in the appendix.

[11] Measuring recidivism from the time of disposition captures incapacitation effects as well as any impacts of criminal behavioral.

V Attorney characteristics

The differences in case outcomes that have been documented above can be the result of several mechanisms. First, attorneys who select to work in a PD office can have different characteristics (e.g., experience, demographics) than those who work as private attorneys and accept appointments from the court (CA). Second, a defendant assigned to a PD office is represented by an organization and not only by a specific attorney. Within the PD office the attorney that is assigned to represent the defendant can consult with other professionals in the office and be exposed to organizational norms and knowledge that have been accumulated through past representation of similar cases.[12]

In the court records of San Francisco, I observe the name of the attorney that represented the defendant and his type (e.g., CA, PD). I use name tabulations from the US Census 2000 and the Social Security Administration to infer the race, ethnicity, and gender of the attorneys from their names. Information on the institutions that awarded B.A. and J.D. degrees to the attorney was obtained using the search engine of the state bar association. To obtain the ranking of each institution I use the information that is publicly available on U.S. News.[13]

Table 1.4 documents the characteristics of PDs and CAs. Relative to CAs, PDs are younger, less experienced, demographically more diverse, and studied in more selective colleges (the best ranking for a university/college is number 1). Two factors that can explain why young individuals who obtained their J.D. in high-ranked universities choose to work in a PD office are (i) ideological motivation and (ii) financial incentives. Regarding the ideological motivation, PDs may desire to represent individuals who cannot afford to hire a private attorney, and will be over-represented by individuals from minority communities compare to the general population. As for financial, Field (2009) documents how in recent years higher ranked J.D. programs provide fee remissions and subsidies to students who work in public interest jobs after graduation. Working in a PD office is considered a public-interest job, unlike being a CA.

To understand how much of the causal attorney type effect can be explained by attorney characteristics, I add them as additional controls to the regression specification in equation (1.2). Including attorney characteristics doubles the standard errors and accounts for different shares of the estimated attorney type effects depending on the case outcome. For prison length, the inclusion of attorney characteristics has almost no impact on coefficient, however, for conviction it explains almost all of the effect (see Appendix Table A.3). Once attorney characteristics are controlled for the estimated differences in case outcomes have the same sign, but become nosier, and not statistically significant.

---

[12] For example, in San Francisco the chief public defender (Jeff Adachi) advocated for the use of checklists by PDs to improve the case outcomes of clients (Adachi, 2015). Writing and disseminating checklists among the attorneys in the office is an example of the advantages of working in an organization that accumulates knowledge and shares it with its members.

[13] U.S. News publishes a ranking of universities and colleges in the U.S. The ranking can be for the entire institution or for a specific program such as a law school. `https://www.usnews.com/`.

This exercise demonstrates that assignment to a PD can impact case outcomes through multiple channel. The first, is that the individual is assigned to an organization with all the resources, case loads, and norms that go with it. Second, the attorneys who work at that organization are different than the ones who acts as CAs. These differences in observed and unobserved characteristics also explain some of the estimated PD assignment effects on case outcomes. Importantly, the characteristics of the attorney are *not* confounders when seeking to identify the effect of assignment to a PD relative to a CA, rather, they are some of the causes/mechanisms that are driving the estimated effects from Section IV.

In addition, representation by an organization is inherently different than being assigned to a specific attorney. Indigent defendants are not assigned to an individual attorney in the PD office by the court, but rather to an organization. The PD office determines how to divide the workload among its attorneys. One attorney can represent the defendant at the initial stages of the case and another at the more advanced court proceedings, including the plea negotiations. Appendix Table A.4 reports the differences in the characteristics of the attorney that first represents the defendant and the terminating attorney. Overall, defendants assigned to the PD office change individual attorneys more often, 52%, relative to 20.7% among those assigned a CA. It is also more frequent that the terminating attorney has more years of experience than the initial attorney among defendants assigned to the PD office.

VI The importance of the attorney type in federal courts

Next I investigate the importance of being assigned a PD relative to a CA in federal district courts. The federal system is statewide and provides estimates that are based on defendants and attorneys from all over the U.S. Thus, this analysis provides evidence on the external validity of the analysis in San Francisco both in terms of the results and the design.

Defendants in federal district courts

Like defendants in San Francisco, the vast majority of federal defendants are represented by indigent defense services. The provision of these services is laid out by the Criminal Justice Act of 1964 and its guidelines are set by the Judicial Conference of the United States.[14] Appendix Table A.5 show descriptive information of defendants in federal district courts for cases terminated between 1996 and 2014. Cases in federal district courts usually end in a conviction and almost always through a plea bargain. Defendants in a multiple defendant case, on average, face more severe charges and their cases terminate with longer prison

[14] The guidelines for the administration of the Criminal Justice Act are described online at: `http://www.uscourts.gov/rules-policies/judiciary-policies/criminal-justice-act-cja-guidelines`

sentences—similar to San Francisco—which puts them in crucial need of legal counsel.[15] It is important to note that I observe detailed demographic and criminal history information only in San Francisco, but not in federal courts; however, in both judicial systems detailed informations on the type and severity of the filled charges is observed.

The number of federal defendants has increased consistently between 1996 and 2014, and the proportion of indigent defendants has increased as well. Appendix Figure A.3 reports the share of defendants represented by a PD, a CA, and/or a privately retained counsel. The share of defendants that are represented by PDs has also increased continuously over time and the share of defendants that retained a private counsel shows a downward trend. In 1970 approximately 40% of all federal defendants retained a private counsel, and in 2014 less than 20% did. The increasing share of defendants who cannot hire a private attorney highlights the importance of studying the provision of indigent defense services.

The number of federal PD organizations has steadily increased since the establishment of the federal PD program in 1970. In 1993, 48 PD organizations served 54 of the 94 federal districts (Prado et al., 1993), and as of 2016, 81 PD organizations have provided indigent defense services to 91 of the 94 federal districts.[16]

Institutional setting

The mechanism in which indigent federal defendants are appointed a legal counsel is different between districts and over time. Prado et al. (1993) describes the CA appointment process: "Some districts have systems in place to ensure an objective rotational system while others base an assignment decision on personal knowledge of an attorney's ability and skill level. In some districts the federal defender office assigns cases; in some districts an employee of the court is given the responsibility." A federal district has broad discretion in how it supplies indigent defense services. 18 U.S. Code § 3006A requires each federal district to prepare an indigent defense plan and to approve it by the judicial council of the circuit (see Chapter 2, § 210.10.10 (d), Appx 2A).[17] The indigent defense plan is obliged to satisfy a list of requirements, one of which is that "private attorneys shall be appointed in a substantial

[15] The federal court records come from through the "Federal Court Cases: Integrated Database", which is constructed by the Bureau of Justice Statistics and made available by the National Archive of Criminal Justice Data and the Inter-University Consortium for Political and Social Research at the University of Michigan. The data series covers every criminal case, in federal district courts, that was terminated from 1970 to 2014. It contains rich information on filed charges, case disposition, and sentencing outcomes. From 1996 onward, both the initial and final attorney types are available. Prior to 1996, only the attorney type at the time of disposition was recorded. For this reason, the main analysis uses only cases that have been terminated/disposed from 1996 onwards.

[16] Although the Judicial Conference called "for the appointment of public defenders in those districts in which the amount of criminal litigation justified the presence of such an office" (Prado et al., 1993), the provision of indigent defense services using PD organizations started only in 1970 and before that time only CAs represented indigent defendants.

[17] See this link for a template of a federal district indigent defense plan, `http://www.uscourts.gov/file/vol07a-ch02-appx2apdf`.

proportion of the cases" (18 U.S Code § 3006A(a)(3)); where "a substantial proportion" is interpreted as 25% of all indigent defense appointments on an annual base (Chapter 2, § 210.10.10 (d), Appx 2A).

Validation tests

I begin by confirming that the conflict of interest consideration cause PD organizations to usually not represent more than one individual within a multiple defendant case. Appendix Figure A.5 clearly validates the conflict-of-interest hypothesis that the PD organization will usually not represent more than one defendant within a multiple defendant case. This result is consistent with the observed patterns in San Francisco. In federal courts similar patterns emerge (Appendix Figure A.4) as is discussed in Section VI.

Overall, defendants who are assigned to a PD organization are significantly different in the charge characteristics as compared to defendants who are assigned a CA. To empirically test for differences in observable characteristics between defendants assigned to a PD relative to a CA, I employ the following econometric model:

$$X_i = \beta \cdot PD_i + \alpha_{j(i)} + \delta_{n(i)} + \xi_i, \tag{1.4}$$

where α_j, and δ_n are case and indictment order of appearance (e.g., first, second, third) fixed effects. This model is analogues to model (1.1) for federal district courts, and the β coefficient can be interpreted as the difference in means in characteristic X_i between defendants who have been assigned a PD relative to a CA.

Appendix Table A.7, columns (1) and (2), shows differences in the severity measures of the filed charges after adjusting for district and filed year fixed effects. The comparisons are based on all indigent defendants (or all multiple defendants) and show that overall defendants who are assigned to a PD face less severe charges relative to defendants who are assigned a CA: a shorter predicted prison term, fewer felony-level charges, a slightly lower predicted probability of a conviction, and a lower predicted probability of a dismissal of charges. The results in columns (1) and (2) are similar to the ones in San Francisco and indicate a similar pattern in both state and federal courts.

The main analysis sample is multiple defendant cases with both a PD and a CA within a case. This sample is different than the overall sample of indigent defendants in two respects: (i) It includes only individuals in multiple defendant cases, and (ii) It restricts attention to multiple defendant cases with both a PD and a CA. Appendix Table A.7 shows that comparisons within multiple defendant cases suffer from the same selection patterns of the overall sample. For example, the cross-case comparison can compare a case with two CAs to one with a PD and a privately retained counsel.

Appendix Table A.7, column (3), shows differences in charge severity between defendants assigned a PD and a CA within a case. This comparison reveals a reverse pattern of selection in the attorney type assignment compared to the cross-case comparison. The PDs are assigned to the defendant facing the more severe charges. In federal courts, the within-case comparison yields a bias estimate of the attorney type effect if the order of the defendants on the indictment is not taken into consideration, and column (3) documents this within-case selection pattern. Appendix Figure A.6 shows the distribution of defendants across attorney types by the position of the defendant on the indictment in multiple defendant cases. This naïve comparison does not reveal differences in the probability of the first defendant on the indictment being assigned a PD relative to a CA. However, as the defendant is further down on the indictment his probability of being assigned a PD decreases dramatically. For example, among defendants who are listed on the first position of the indictment the share who are assigned a PD is approximately 40% relative to approximately 10% who are assigned a PD among defendants listed on the third position of the indictment. The position of the defendant on the indictment is a strong predictor of the attorney type that will be assigned to the defendant.

To examine whether the order of the defendants on the indictment is correlated with sentencing outcomes of interest such as length of incarceration I estimate the following model:

$$\text{asinh}(\text{Prison term})_i = \delta_{p(i)} + \alpha_{c(i)} + \kappa_i \tag{1.5}$$

As length of incarceration is an extremely skewed distribution a common practice is to perform some concave transformation (e.g., a logarithmic function). When the outcome of interest has a point mass at zero the asinh$(\cdot)$ function is commonly used as an approximation for the logarithmic function.[18] Appendix Figure A.7 plots the estimated $\delta_{p(i)}$ coefficients and presents compelling evidence that the order on the indictment is a strong predictor of the length of incarceration that the defendant will be sentenced. The first defendant is likely to face a harsher sentence than the other defendants listed on the indictment. The order of defendants on the indictment can be considered as an additional measure of the differences in unobservable confounders (e.g., culpability) between the defendants.

To take into account the order of the defendants on the indictment, I estimates the model in equation (1.4) that includes a specific fixed effect for each position on the indictment. Appendix Table A.7, column (4), shows that after conditioning on the defendant's order of appearance on the indictment there are *no* differences in the charge severity measures between defendants assigned a PD relative to a CA.

Figure 1.6 provides a visualized summary of the balance tests reported in Appendix Table A.7. The figure illustrates both the non-random sorting that is present in a naïve comparison and the comparability of defendants assigned to different attorney types within a case after conditioning on the position on the indictment.

As an additional balance test, I ranked the defendants within each case by their predicted prison term (months) and defined an indicator variable for the defendant who faced the highest predicted prison term within a case. In the same way I rank defendants within a case based on other predicted outcomes such as the probability of being convicted or the

[18]For example, see Gelber, Isen, , and Song (2016) who apply this approximation to Social Security Administration earning records or Card and DellaVigna (2013) who apply it to citations of academic papers, which is also a skewed distribution with a large mass at zero.

probability that a trial will take place. Appendix Table A.8 reports the difference in means in the probability that the PD organization will be assigned to the highest-ranked defendant. Each row in the table refers to a ranking based on a different predicted sentencing outcome, and each cell is a coefficient from a different regression specification. The PD is more likely to be assigned to the more severe defendant within a case; however, this bias disappears once conditioning on the defendant's position on the indictment.

Attorney type effects

To estimate the effect of being assigned a PD relative to a CA in federal courts, I consider a second causal model that flexibly controls for both case level factors and the order of appearance of the defendants on the indictment:

$$Y_i = \beta \cdot \mathrm{PD}_i + X_i'\Gamma + \alpha_{j(i)} + \delta_{n(i)} + \epsilon_i \tag{1.6}$$

where $\delta_{n(i)}$ is a fixed effect for each position on the indictment (e.g., first, second) and $n(i)$ is a mapping from defendant i to his position on the indictment n. The identifying assumption is that once we condition on the defendant's position on the indictment there are no unobserved confounders that are correlated with both the defendant's potential outcomes (i.e., culpability) and the assignment to a PD. Section VI, documents that in federal district courts, unlike San Francisco, there is within-case selection that can confound a causal interpretation of model (1.2). However, after conditioning on the defendant's position on the indictment, balance tests on observable characteristics suggest that estimation via OLS recovers a causal relationship.

In federal courts, to estimate the attorney type causal effect it is also necessary to condition on the order of appearance of the defendants on the indictment (i.e., model (1.6)). Table 1.5 shows the estimation results and it also highlights the selection challenges that must be overcome to estimate the causal effect of attorney type on defendant's sentencing outcomes. The unadjusted asinh(Prison term) estimate in column (1), -0.278, is similar to the estimate in San Francisco, -35.7, and after adjusting for the charge codes the coefficient shrinks to -0.0322. Unlike San Francisco, in federal courts the order of the defendants on the indictment has a large impact on the estimates and a simple within-case comparison yields a coefficient of 0.115 after covariate adjustment. The within-case comparison in federal courts yields an opposite result to the estimates in San Francisco. However, after controlling for the defendant's position on the indictment (columns (7) and (8)), the estimated effect, -0.0462, has the same sign as the one in San Francisco, -0.109.

According to Table 1.5, PDs obtain shorter prison sentences for their clients (4.64%), a slightly higher probation term (2.39%), and a lower probability of any prison term (0.819pp). I find no differences in the probability of reaching a plea bargain on some of the charges; however, I find a 0.768pp statistically significant lower probability of taking a case to trial. This is a small estimated coefficient, but relative to the average number of cases that go to trial (5% in this sample) it implies a 15.6% decrease in the probability that a trial will take place. I find no differences in the probability of conviction or acquittal of defendants initially assigned a PD relative to a CA.

Taken together, the findings from both San Francisco and federal district court present empirical evidence that the method by which indigent defense services are provided, PD vs. CA, influences the trial outcomes of the defendants. Indigent defendants in multiple defendant cases who were assigned a PD obtained more favorable outcomes than their co-defendants who were represented by a CA.

VII Discussion

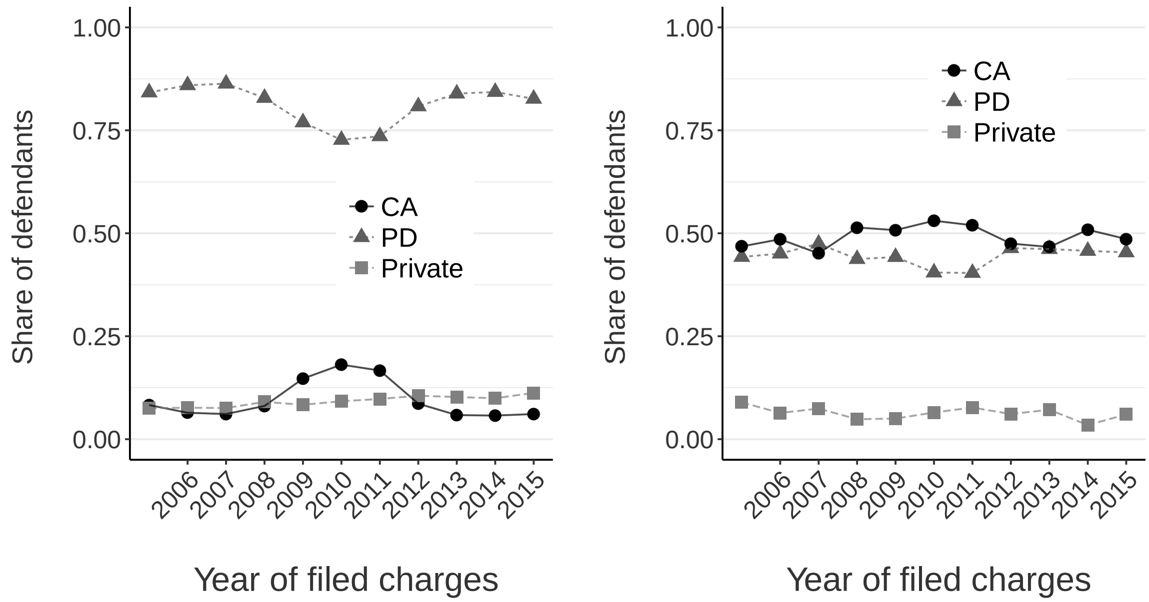
The vast majority of defendants facing criminal charges require the assistance of court-appointed legal counsel. This paper develops a framework to compare two methods of providing legal representation to defendants who cannot afford to hire an attorney in the private market. I use a new empirical identification strategy and administrative court records from both state and federal district courts to study whether defendants assigned to a public defender (PD) obtain better or worse case outcomes than those who are represented by a private court-appointed attorney (CA). The results have direct policy implications on how indigent defense representation should be provided and whether the current system violates defendants' Sixth Amendment rights.

Defendants who have been represented by a PD, relative to a CA, obtained more favorable case outcomes (e.g., shorter prison sentences, lower probability of any imprisonment). In San Francisco, defendants who face more severe charges (felony vs. misdemeanor) and have a longer criminal history are the ones driving the results. Those who face a higher risk of imprisonment are the ones on whom the attorney type makes the largest impact. One explanation for these differences is that attorneys who work for a PD organization are substantially different in their observable characteristics from those who self-select to act as CAs. PDs have fewer years of experience, are demographically more diverse, and studied in more selective institutions both in their B.A. and J.D. programs.

The method of provision of indigent defense services is part of the bigger question of how the state should supply public services (e.g., police, prison). Should the state establish a PD organization or use the private sector and hire CAs? To answer this question, one of the key issues that needs to be addressed is which kind of attorney will select to represent low-income defendants under each one of the aforementioned alternatives. Future research is needed to examine what motivates attorneys to select to work in a PD office, relative to the self-selection of those who act as CAs. More information is needed to understand how policy makers can mitigate the attorney type differences that are documented in this study.

Figures

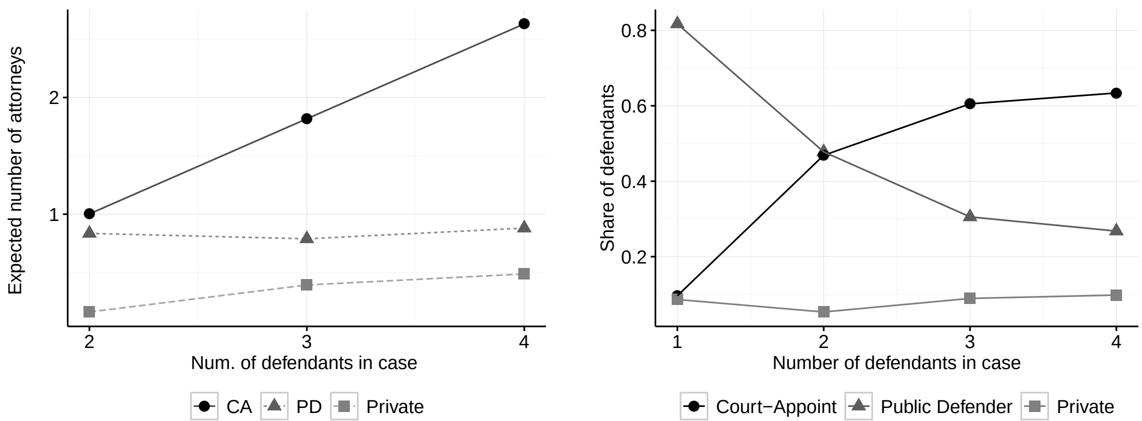
Figure 1.1: San Francisco: The distribution of defendants across attorney types and over time, by filing year



A. Single defendant B. Multiple defendants

Notes: The figure shows the distribution of criminal defendants in San Francisco across attorney types. The left plot shows the distribution of defendants across attorney types in cases with a single defendant. The right plot shows the distribution of defendants across attorney types in cases with multiple defendants.

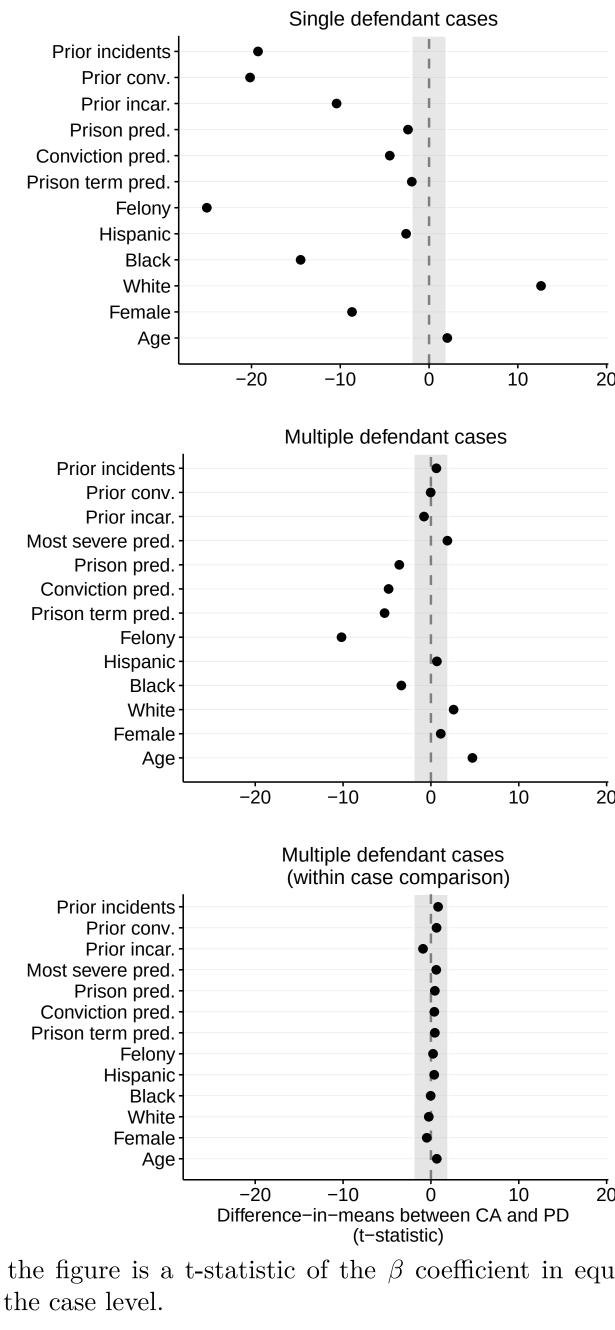
Figure 1.2: Validating the conflict of interest hypothesis



A. Ave. num. of attorneys by type B. Distribution of across attorneys

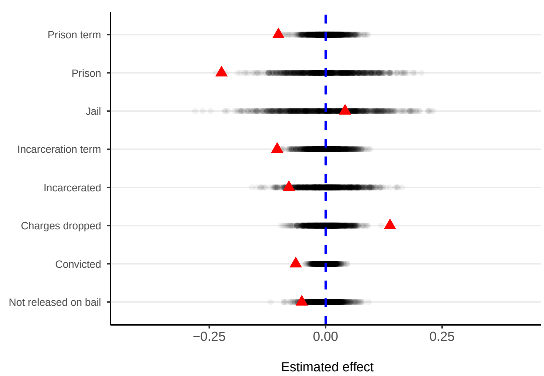
Notes: The figure presents descriptive evidence on the distribution of defendants across attorney types in multiple defendant cases in San Francisco. The left plot (Panel (a)), the x-axis describes the number of defendants in a case and the y-axis the average number of attorneys from each type (e.g. PD or CA). For example, Panel (a) shows that in multiple defendant cases in San Francisco with three defendants there are on average almost 2 CAs and approximately one PD. Panel (b) shows the distribution of attorney types (i.e., the share of defendants represented by PD, CA) by the size of the multiple defendant case. As the number of co-defendants in a case increases the share who are assigned to a PD decreases and the share who are assigned a CA or represented by a private attorney increases.

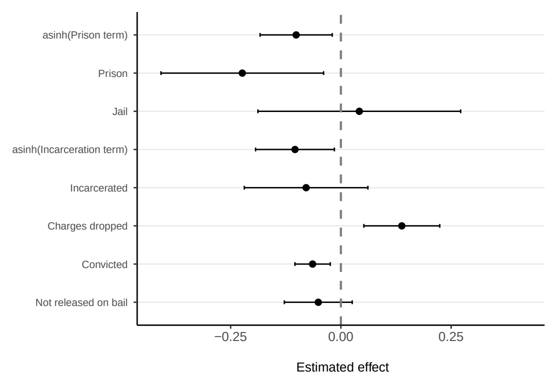
Figure 1.3: San Francisco: Differences in observable characteristics between defendants assigned to PD vs. CA (2006 – 2016)



Notes: Each point on the figure is a t-statistic of the β coefficient in equation (1.1). Standard errors are clustered at the case level.

Figure 1.4: San Francisco: The effects of the attorney type on the defendant's case outcomes

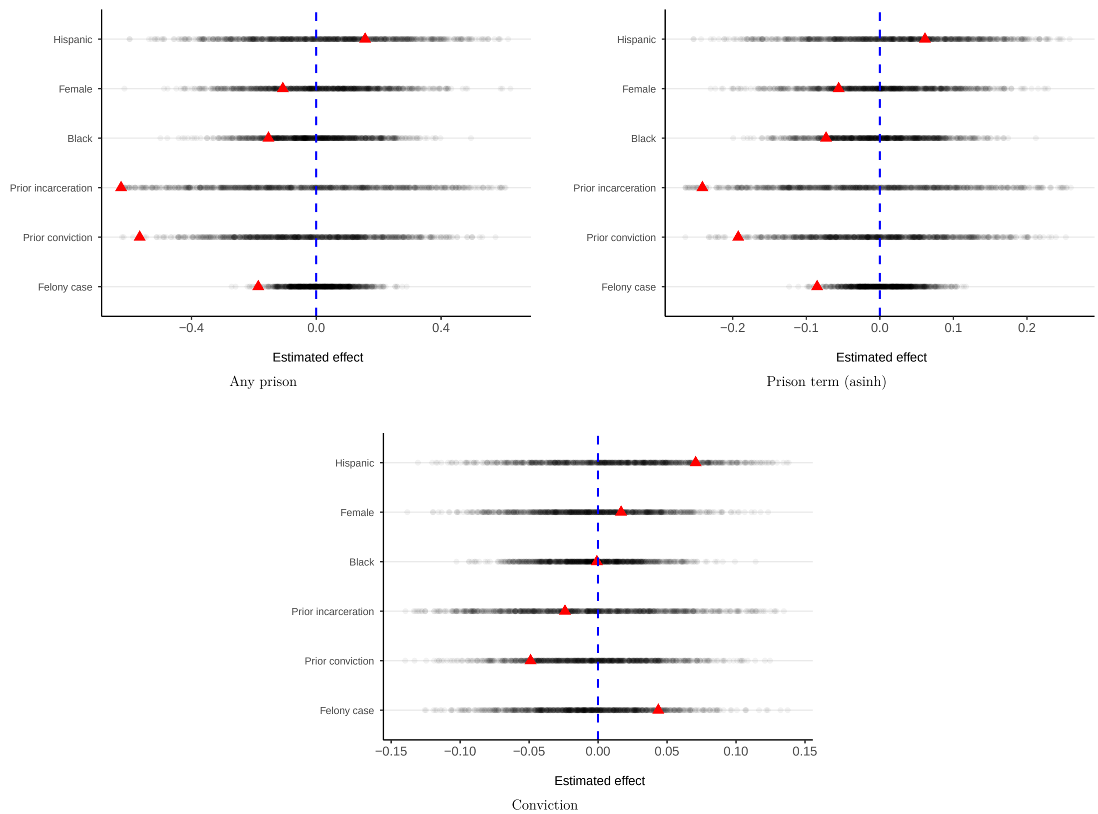




A. Permutation inference B. Cluster-robust standard errors

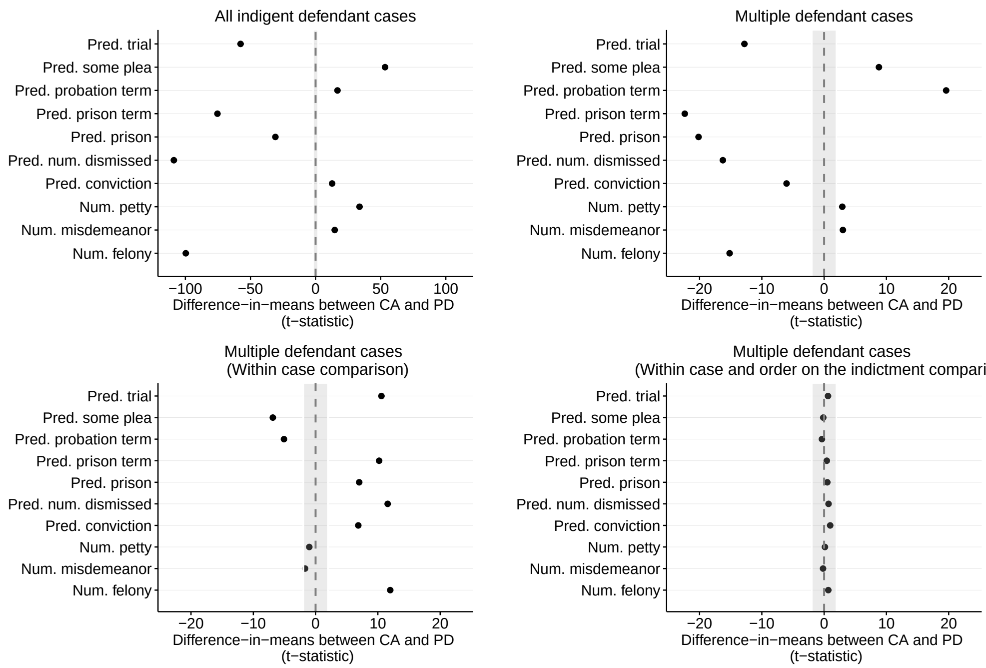
Notes: Confidence intervals for selected coefficients from Table 1.3 are presented divided by the baseline mean value of the sentencing outcome of interest. As in Table 1.3, standard errors are clustered at the case level. The right plot displays 1,000 estimates of the β coefficient from equation (1.2), where defendants in the same case have been randomly labelled as being represented by a PD—the randomization is within a case. The black dots are from each one of the permutations. Areas which are darker indicate a higher chance of observing those values due to random chance, when the attorney type has no impact on the case outcomes. The red triangle indicate the observed values in the data, which are not likely under the null that the attorney type has no effect on the case outcomes. The number of re-labellings we use is 1,000 and it is similar to what is commonly used in the statistics literature. For example, Athey, Eckles, and Imbens (2018) and Anderson and Magruder (2017) use 1,000 draws/re-labellings; and Keele and Miratrix (2017) use 500 draws/re-labellings. The Log($\cdot$) function is approximated using the arcsinh function which is commonly used as an approximation to the logarithmic function when the variable of interest contains values of zero.

Figure 1.5: San Francisco: Heterogeneity in the effects of the attorney type on the defendant's case outcomes by criminal history and demographic characteristics



Notes: See the notes to figure (1.4). The only difference is that here the estimates are not of the *beta* coefficient from equation (1.2), but rather the γ coefficient from equation (1.3) for different dimensions of possible heterogeneity (e.g., gender, criminal history). Each one of the plots is for a different sentencing outcome (e.g., any prison sentence, conviction).

Figure 1.6: Federal courts: Differences in observable characteristics between defendants assigned to PD vs. CA (1996 – 2014)



Notes: Each point on the figure is a t-statistic of the β coefficient from model specification (1.4). Standard errors are clustered at the case level. The two upper plots show the results from specifications without case FE (but with district and calendar year FEs), and the bottom figures reports the results once case FE are included. The bottom right (left) figure includes (does not include) controls for the defendant's position on the indictment. The standard errors are cluster-robust at the case level. The gray area represents the 95% confidence interval in which the null that the coefficient β is zero cannot be rejected. Since the number of observations when estimating a given specification for each of the outcomes is the same; and the figure reports t-statistics, rather than β coefficients, then the gray area is the same for all the t-statistics of each of the outcomes and is approximately ± 1.964 around zero.

Tables

Table 1.1: San Francisco, defendants' characteristics in single and multiple defendant cases

	Single defendant (1)	Multiple defendant (2)	Multiple design (3)
Age	35.503	32.250	32.181
White	0.530	0.453	0.435
Black	0.440	0.515	0.535
Female	0.177	0.239	0.238
Hispanic	0.195	0.200	0.201
Highest filed charge felony	0.518	0.829	0.873
Predicted conviction	0.586	0.614	0.619
Predicted prison>0	0.057	0.058	0.058
Num. prior incarcerations	0.174	0.146	0.158
Num. prior convictions	0.521	0.470	0.485
Num. prior arrests	2.237	2.027	2.113
Dropped charges	0.246	0.249	0.230
Convicted	0.598	0.586	0.609
Prison	0.063	0.080	0.079
Jail	0.098	0.070	0.071
Observations	64,191	9,576	7,164

Notes: The table presents descriptive statistics for all criminal defendants in San Francisco between 2006 and 2016. Columns (1) and (2) include all incident-defendant pairs in the analysis data set and are not restricted to indigent defendants. Columns (3) includes only indigent defendants in multiple defendant cases which have both a PD and a CA. The third columns reports the descriptive statistics for the main analysis sample in which the assignment of defendants to attorney type is as good as random within a case.

Table 1.2: San Francisco: Differences in observable characteristics between defendants who are assigned PD and CA

	All indigent	All Multiple	Multiple PD & CA		Co-defendant PD & CA	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	1.624***	1.064***	0.630**	0.158	0.225	0.225
	(0.124)	(0.252)	(0.280)	(0.292)	(0.312)	(0.337)
Female	−0.050***	0.010	−0.002	−0.006	−0.007	−0.007
	(0.004)	(0.009)	(0.010)	(0.013)	(0.011)	(0.015)
White	0.084***	0.022**	0.004	0.002	−0.003	−0.003
	(0.005)	(0.010)	(0.012)	(0.011)	(0.013)	(0.013)
Black	−0.091***	−0.029***	−0.004	−0.006	−0.001	−0.001
	(0.005)	(0.011)	(0.012)	(0.011)	(0.013)	(0.012)
Hispanic	−0.010**	0.005	0.002	0.001	0.007	0.007
	(0.004)	(0.008)	(0.009)	(0.013)	(0.011)	(0.015)
Felony	−0.220***	−0.069***	−0.007	0.001	0.001	0.001
	(0.005)	(0.008)	(0.008)	(0.003)	(0.009)	(0.003)
Predicted prison term	−0.771***	−0.783***	−0.170	0.008	0.089	0.089
	(0.090)	(0.162)	(0.178)	(0.180)	(0.175)	(0.203)
Predicted convicted	−0.018***	−0.009***	−0.0001	0.001	0.001	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Predicted prison	−0.002***	−0.002***	−0.00000	0.0001	0.0004	0.0004
	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Most severe	–	0.023**	0.025**	0.008	0.015	0.015
	(0.003)	(0.011)	(0.012)	(0.022)	(0.013)	(0.026)
Num. prior incarceration	−0.053***	−0.008	−0.001	−0.010	−0.018	−0.018
	(0.007)	(0.011)	(0.013)	(0.017)	(0.015)	(0.019)
Num. prior convictions	−0.256***	−0.001	0.056**	0.037	0.025	0.025
	(0.015)	(0.025)	(0.028)	(0.034)	(0.032)	(0.039)
Num. prior incidents	−0.874***	0.053	0.193*	0.101	0.115	0.115
	(0.054)	(0.093)	(0.104)	(0.119)	(0.121)	(0.141)
Observations	67,620	8,975	7,164	7,164	5,826	5,826
Case FE	No	No	No	Yes	No	Yes

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD. The table reports the estimates of the β coefficient from model (1.1). Standard errors are clustered-robust at the case level. Columns 3 and 4 include all multiple defendant cases with both a PD and a CA within each case. Columns 5-6 include only multiple defendant cases with exactly two indigent defendants that one was assigned a PD and the other a CA. For example, a case with 3 indigent defendants that two of which are represented by CAs and the third by a PD will be included in columns 3 and 4 but not in columns 5 and 6.
*p<0.1; **p<0.05; ***p<0.01

Table 1.3: San Francisco, the effect of being initially assigned a PD vs. a CA on the case sentencing outcomes

	coefficient of interest Initial PD indicator					
	All indigent		All Multiple		Multiple PD & CA	
	(1)	(2)	(3)	(4)	(5)	(6)
Convicted	-0.070^{***}	-0.016^{***}	-0.029^{***}	-0.037^{***}	-0.039^{***}	-0.039^{***}
	(0.005)	(0.005)	(0.009)	(0.012)	(0.012)	(0.012)
Ave.	0.592	0.592	0.586	0.609	0.609	0.609
Incarcerate	-0.050^{***}	-0.020^{***}	-0.009	-0.012	-0.014	-0.012
	(0.004)	(0.004)	(0.007)	(0.010)	(0.010)	(0.010)
Ave.	0.162	0.162	0.144	0.146	0.146	0.146
Prison	-0.059^{***}	-0.032^{***}	-0.016^{***}	-0.018^{**}	-0.019^{***}	-0.018^{**}
	(0.003)	(0.003)	(0.005)	(0.007)	(0.007)	(0.007)
Ave.	0.068	0.068	0.08	0.079	0.079	0.079
Jail	0.005	0.010^{***}	0.004	0.003	0.003	0.003
	(0.003)	(0.003)	(0.005)	(0.008)	(0.008)	(0.008)
Ave.	0.099	0.099	0.07	0.071	0.071	0.071
asinh(Incarceration term)	-0.357^{***}	-0.180^{***}	-0.099^{***}	-0.108^{**}	-0.109^{**}	-0.104^{**}
	(0.020)	(0.019)	(0.031)	(0.045)	(0.045)	(0.046)
Ave.	0.546	0.546	0.59	0.597	0.597	0.597
asinh(Prison term)	-0.331^{***}	-0.189^{***}	-0.100^{***}	-0.105^{**}	-0.107^{**}	-0.101^{**}
	(0.019)	(0.018)	(0.028)	(0.041)	(0.042)	(0.042)
Ave.	0.359	0.359	0.436	0.435	0.435	0.435
No bail	-0.086^{***}	-0.019^{***}	-0.010	-0.015	-0.018	-0.018
	(0.005)	(0.005)	(0.009)	(0.014)	(0.014)	(0.014)
Ave.	0.338	0.338	0.342	0.359	0.359	0.359
Case FE	No	No	No	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
Prior PD control	No	No	No	No	No	Yes
Observations	67,613	67,613	9,576	7,164	7,164	7,164

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD or a CA. The standard errors are cluster-robust at the case level, which is the level in which treatment—attorney type—is assigned. Both incarceration and prison terms are measured in months. I approximate the Log(·) function using the asinh(·) function which is a common procedure when the outcome of interest is both skewed and has a mass at zero.

*p<0.1; **p<0.05; ***p<0.01

Table 1.4: San Francisco: Attorney characteristics by attorney type, at the defendant level

	CA	PD	Private
Female	0.253	0.472	0.194
Asian	0.039	0.166	0.036
White	0.573	0.408	0.552
Hispanic	0.031	0.099	0.068
Ave. Rank BA (USnews)	54.661	44.767	51.675
Ave. Rank JD (USnews)	77.015	47.981	65.995
Ave. No rank BA (USnews)	0.169	0.137	0.198
Ave. No rank JD (USnews)	0.895	0.836	0.847
Experience (median)	22.287	6.256	16.776
Num. cases first attorney (median)	39	208	7
Num. cases terminating attorney (median)	56	173	10

Notes: The table shows the characteristics of the initial attorney that represented each defendant. All the calculations in the table where done at the defendant level. The numbers are attorney characteristics averaged across defendants. This is equivalent to the average of attorney characteristics re-weighted by the number of defendants that each individual attorney represented. The “Num. cases first attorney” is the number of cases in which the attorney was the first assigned attorney in a case, and similarly “Num. cases first attorney” is the number of cases in which the attorney was the terminating attorney.

Table 1.5: Federal courts, the effect of being initially assigned a PD vs. a CA on the case sentencing outcomes

	All indigent		All multiple		Multiple (PD & CA)		Multiple (PD & CA)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
asinh(Prison term)	-0.278***	-0.0322***	-0.0841***	0.0368***	0.145***	0.115***	-0.0459***	-0.0464***
	(0.00556)	(0.00453)	(0.0106)	(0.00927)	(0.0122)	(0.0119)	(0.0136)	(0.0134)
Prison	-0.00801***	0.00411***	-0.00435*	0.00788***	0.0166***	0.0132***	-0.00811**	-0.00819**
	(0.00100)	(0.000930)	(0.00197)	(0.00189)	(0.00253)	(0.00252)	(0.00288)	(0.00286)
asinh(Probation term)	0.0237***	-0.00522	0.0433***	0.00340	-0.0170*	-0.0113	0.0246**	0.0239*
	(0.00332)	(0.00314)	(0.00640)	(0.00609)	(0.00825)	(0.00821)	(0.00952)	(0.00944)
Some plea	0.0161***	0.00983***	0.0174***	0.0171***	0.0157***	0.0151***	0.00424	0.00432
	(0.000781)	(0.000754)	(0.00171)	(0.00167)	(0.00221)	(0.00221)	(0.00253)	(0.00253)
Trial	-0.0151***	-0.00787***	-0.0136***	-0.0113***	-0.00652***	-0.00782***	-0.00738***	-0.00762***
	(0.000537)	(0.000506)	(0.00123)	(0.00119)	(0.00155)	(0.00155)	(0.00177)	(0.00177)
Num. conviction	-0.0526***	0.0142***	0.0325***	0.0451***	0.0522***	0.0319***	-0.00390	-0.00502
	(0.00182)	(0.00160)	(0.00428)	(0.00392)	(0.00521)	(0.00490)	(0.00586)	(0.00551)
Conviction	0.00282***	0.00284***	0.00642***	0.00825***	0.0113***	0.00959***	-0.00219	-0.00233
	(0.000637)	(0.000627)	(0.00135)	(0.00133)	(0.00179)	(0.00179)	(0.00204)	(0.00203)
Num. dismissed	-0.365***	-0.0229***	-0.136***	-0.0421***	0.0206**	-0.0199***	0.0135	0.0107
	(0.00405)	(0.00195)	(0.00765)	(0.00478)	(0.00659)	(0.00560)	(0.00745)	(0.00632)
Acquittal	-0.00106***	-0.000780***	-0.00191***	-0.00198***	-0.00197***	-0.00184**	-0.000554	-0.000527
	(0.000203)	(0.000195)	(0.000450)	(0.000442)	(0.000584)	(0.000583)	(0.000646)	(0.000646)
Some diversion	0.00120***	0.000537***	0.000224	0.0000213	-0.000347	-0.000273	-0.000147	-0.000106
	(0.000109)	(0.000108)	(0.000179)	(0.000174)	(0.000243)	(0.000240)	(0.000235)	(0.000234)
N	651,666	651,666	182,875	182,875	84,260	84,260	84,260	84,260
Def. Num. FE	No	No	No	No	No	No	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Case FE	No	No	No	No	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	No	No	No	No

Notes: *p<0.1; **p<0.05; ***p<0.01

Chapter 2

Does Incarceration Increase Crime?

I 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, Dahl, Loken, and Mogstad, 2018b) or deter (Becker, 1968; Drago, Galbiati, and Vertova, 2009) convicted individuals, but it can also serve as a "crime school" by exposing them to criminal peers (Bayer, Hjalmarsson, and Pozen, 2009; Stevenson, 2017). Moreover, the stigma attached to incarceration might disconnect individuals from the labor market once they are released, further increasing criminal behavior (Kling, 2006; Steven, 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 not correlated with individuals' unobserved criminality. To isolate such variation, we use discontinuities in North Carolina's sentencing guidelines. These guidelines set allowable sentences as a function of individuals' offense severity and criminal history, which is aggregated into a numerical score that counts prior convictions weighted by seriousness. Guideline sentences change discretely when offenders' criminal history scores cross critical thresholds, providing shifts in the sentence type (incarceration vs. probation) and length for otherwise comparable individuals. For example, the likelihood of incarceration increases by more than 30 p.p. between 4 and 5 criminal history points for offenders convicted of first degree burglary. We focus on 5 discontinuities that shift both the likelihood of being incarcerated and the length of incarceration, but also explore 15 others that primarily shift the length of incarceration.

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 for a period of time 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 year period after sentencing. This negative (i.e., crime reducing) effect persists over longer windows and is still evident even eight years after sentencing.

At least part of this effect is directly attributable to incapacitation. To explore the dynamic evolution of the effects, 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 in a given month 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 and monthly offending rates for the two groups are indistinguishable. However, initial incarceration still causes a reduction in cumulative measures of crime such as *ever* reoffending in the eight years after sentencing.

To further explore the role of incapacitation vs. other behavioral effects of incarceration, 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, behavioral effects are directly estimable (Nagin and Snodgrass, 2013; Mears, Cochran, Bales, and Bhati, 2016; Harding, Morenoff, Nguyen, and Bushway, 2017). We find no evidence of any criminogenic impacts. An additional year of incarceration reduces reoffending within three years of at-risk by either 8.9 p.p. (↓19%) or 0.46 p.p. (↓1%), depending on how reoffending is measured.

Thus far, our analysis has been limited by several key factors. First, 2SLS estimates from conviction and from at-risk are parameterized by a single continuous endogenous variable (months of incarceration) 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 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 sentencing discontinuities and 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 show below, 2SLS estimates based on measuring reoffending from at-risk recover a mixture of the behavioral effects of incarceration (e.g., exposure to criminal peers) and the effects of other time-varying factors (e.g., aging, unemployment rates, etc.).

In the second part of our study, we present a unified framework that overcomes these core challenges and allows us to unpack and interpret the reduced form evidence. We develop a semi-parametric model of incarceration length and recidivism that describes how the latent propensity to commit crime varies with incarceration length and release date. This framework 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, a single index ordered-choice model of assignment to different lengths of incarceration is estimated via maximum likelihood. The generalized residuals (Gourieroux, Monfort, Renault, and Trognon, 1987) are used to proxy for the latent criminality that generates omitted variable bias in the assignment to incarceration. Second, we eliminate the initial period of incapacitation by resetting the starting point at which reoffending is measured to each individual's at-risk date. To isolate the non-linear behavioral effects of incarceration from any time-varying factors, we model the likelihood of reoffending within t months from at-risk as a function of time-invariant covariates (e.g., criminal history), time-varying factors (e.g., age at release), a control function, and a flexible function of incarceration length. The behavioral effects are then estimated using a series of OLS regressions of the relationship between incarceration length and reoffending within t months from at-risk. Incapacitation effects are given by model estimates of changes in offending rates due to changes in time at-risk. 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 the effects of incarceration 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, 2007a,b). 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, Ham, and Lalonde, 1997) or military service.

The model estimates reveal 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 policies. 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, a marginal increase of 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 due to more aggressive sentencing (through both incapacitation and behavioral channels), the high cost of incarceration likely outweigh 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. (2018b) 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

[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) are Kling (2006), Green and Winik (2010), Loeffer (2013), Nagin and Snodgrass (2013), Mueller-Smith (2015), Aizer and Doyle (2015a), Stevenson (2016), Harding et al. (2017), Zapryanova (2017), Arnold et al. (2018),Arteaga (2018), Aneja and Avenancio-León (2018), Bhuller et al. (2018b), Bhuller, Dahl, Loken, and Mogstad (2018a), Dobbie, Goldin, and Yang (2018a), Dobbie, Grönqvist, Niknami, Palme, and Priks (2018b) , Norris (2018), and Norris, Pecenco, and Weave (2018).

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. (2018b) and differ in both in sign and magnitude from Mueller-Smith (2015). This may reflect differences in the causal 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. (2018b) 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, Harding, Morenoff, and Bushway (2017) find that reincarceration rates are higher for initially incarcerated offenders. Differences in the institutional setting and whether incarcerated or non-incarcerated offenders are differentially impacted by technical revocations of probation or parole can potentially explain some of the differences across these studies. 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. Estimates from this literature also vary widely (Levitt, 1996; Raphael and Lofstrom, 2015).

[3] Studies on juvenile offenders also find mixed results (Aizer and Doyle, 2015a; Hjalmarsson, 2009), however, the effects of incarceration may be substantially different for juvenile vs. 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).

II 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, Mogstad, and Wiswall, 2012; Lochner and Moretti, 2015) and heterogeneous effects.

Two approaches to measuring reoffending

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 involved in crime (e.g., arrested or charged with a new offense) 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. The probability of reoffending within a certain time window from the date of conviction/release constitutes a “failure function" in the terminology of duration analysis. A key question is when to start measuring time until a new offense. 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 (i.e., is not in jail or prison) and is therefore at-risk to reoffend.

Notation

We present a unified potential outcomes framework that formalizes what causal parameters are identified under each approach when using instrumental variables to overcome non-random assignment. To do so, we model failure functions as potential outcomes. Let $Y_{i,t}(d)$ denote 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 is 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: $Y_{i,t+d}(d)$. In our setting, the instruments are indicators for being above a discontinuity (described in full detail below), which we denote $Z_i \in \{0, 1\}$. Let $D_i(z)$ denote the number of months individual i was initially incarcerated as a function of whether she is to the left ($z = 0$) or to the right ($z = 1$) of a punishment discontinuity. The realized incarceration length can be expressed as $D_i = D_i(1)Z_i + D_i(0)(1 - Z_i)$ and observed reoffending can be written as $Y_{i,t} = Y_{i,t}(D_i(1))Z_i + Y_{i,t}(D_i(0))(1 - Z_i)$.

Identification using an IV

In this section, we describe what estimands are identified by different IV estimators. This illustrates how to interpret our estimates and to motivates our choices among the different possible IV estimators. The discussion below is relevant for any setting with a binary instrument and an ordered treatment with multiple levels (e.g., years of education).

The IV approach accounts for the fact that incarceration assignment is unlikely to be independent of individuals' propensity to reoffend. The instrument is assumed to satisfy the usual assumptions of the LATE framework (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 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 in their characteristics; the only difference lies in their exposure to incarceration sentences. Assumption 3 states that being above a punishment discontinuity weakly increases the sentences of all offenders. It rules out a scenario where some individuals face more lenient sentencing to the right of a discontinuity, while others do not. We provide a direct test of the monotonicity assumption in Section IV.

A complier in our setting is an individual that is incarcerated for a longer duration because he or she is located to the right of a punishment discontinuity, i.e., $D_i(1) > D_i(0)$. Since incarceration length is a treatment with multiple levels there are many types of compliers that represent different changes in the exposure to incarceration. A complier of type d is an individual who was 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, $\mathbb{I}\{D_i > 0\}$. When using D_i as the endogenous variable, IV recovers the "average causal response" discussed in Angrist and Imbens (1995):

$$\beta_{\text{conviction}}(t) \equiv \frac{\mathbb{E}[Y_{i,t}|Z_i = 1] - \mathbb{E}[Y_{i,t}|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}\left[Y_{i,t}(d) - Y_{i,t}(d-1) | \underbrace{D_i(1) \geq d > D_i(0)}_{\text{Compliers of type } d}\right] \tag{2.1}$$

where

$$\omega_d = \frac{\Pr(D_i(1) \geq d > D_i(0))}{\sum_{l=1}^{\bar{D}} \Pr(D_i(1) \geq l > D_i(0))} \tag{2.2}$$

Equation (2.1) shows that the IV estimand $\beta_{\text{conviction}}(t)$ is a weighted average of causal effects for different populations of compliers. E.g., $\mathbb{E}\left[Y_{i,t}(d) - Y_{i,t}(d-1) \mid D_i(1) \geq d > D_i(0)\right]$ is the effect of an additional month of incarceration on the likelihood of reoffending within t months of conviction for individuals that would have been incarcerated for strictly less than d months without the instrument ($Z_i = 0$) and that if assigned to the instrument ($Z_i = 1$) would be incarcerated for at least d months. Although $\beta_{\text{conviction}}(t)$ recovers a combination of incapacitation and behavioral effects, this composite impact is a relevant policy parameter when evaluating the overall effect of incarceration length on reoffending in the years after conviction.[5]

In an attempt to separate incapacitation effects from any behavioral effects of incarceration post 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

$$\begin{aligned}\beta_{\text{at-risk}}(t) &\equiv \frac{\mathbb{E}\left[Y_{i,t+D_i}|Z_i = 1\right] - \mathbb{E}\left[Y_{i,t+D_i}|Z_i = 0\right]}{\mathbb{E}\left[D_i|Z_i = 1\right] - \mathbb{E}\left[D_i|Z_i = 0\right]} \\ &= \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}\left[Y_{i,t+d}(d) - Y_{i,t+d-1}(d-1)|D_i(1) \geq d > D_i(0)\right]\end{aligned} \tag{2.3}$$

A sketch of the proof of Equation (2.3) is presented in Appendix II. The estimand in Equation (2.3) 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 can directly influence offenders' criminal propensities. Without imposing additional structure on the potential outcomes (e.g., additive separability and a constant linear treatment effect) it is not possible to express treatment effect estimates as a behavioral effect and a bias term due to aging (or other time varying factors). In other words, there is no non-parametric potential outcomes representation of the behavioral effect because it does not correspond to a well-defined hypothetical manipulation. In Appendix II, we present an example for an explicit model of potential outcomes, which allows us to articulate exactly the behavioral effects separately from time-varying factors. Moreover, the example also illustrates how the 2SLS estimates can be represented as recovering a mixture of the effects of behavioral responses and time-varying factors.

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 in the IV regression can potentially lead to bias since these variables are functions of incarceration length, and therefore endogenous. In Appendix II, we present an example of how controlling for such time-varying factors

[5] In Appendix I, 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.

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 of the selection model based analysis in Section 3.

Finally, researchers often use a binary indicator for whether the individual was incarcerated or not, $\mathbb{I}\{D_i > 0\}$, as the endogenous treatment. Angrist and Imbens (1995) showed that:

$$\gamma_{\text{conviction}}(t) \equiv \frac{\mathbb{E}[Y_{i,t}|Z_i = 1] - \mathbb{E}[Y_{i,t}|Z_i = 0]}{\mathbb{E}[1(D_i > 0)|Z_i = 1] - \mathbb{E}[1(D_i > 0)|Z_i = 0]} = \beta_{\text{conviction}}(t) \cdot (1 + \kappa) \tag{2.4}$$

where

$$\kappa \equiv \frac{\sum_{l=2}^{\bar{D}} \Pr(D_i(1) \geq l > D_i(0))}{\Pr(D_i(1) \geq 1 > D_i(0))} \tag{2.5}$$

and therefore using $\mathbb{I}\{D_i > 0\}$ as the endogenous treatment yields a biased estimate of the average causal response $\beta_{\text{conviction}}(t)$; however, $\gamma_{\text{conviction}}(t)$ can still have a causal interpretation as capturing a different treatment effect than the average causal response. Specifically, Equation (2.6) 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 ($D_i(1) > D_i(0) > 0$).

$$\begin{aligned} \gamma_{\text{conviction}}(t) = {} & \underbrace{\mathbb{E}[Y_{i,t}(D_i(1)) - Y_{i,t}(0)|D_i(1) > D_i(0) = 0]}_{\text{Extensive margin}} + \\ & \left(\underbrace{\mathbb{E}[Y_{i,t}(D_i(1)) - Y_{i,t}(D_i(0))|D_i(1) > D_i(0) > 0]}_{\text{Intensive margin}} \right) \frac{\Pr(D_i(1) > D_i(0) > 0)}{\Pr(D_i(1) > D_i(0) = 0)} \end{aligned} \tag{2.6}$$

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.[6]

[6]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.

To provide interpretable and well-defined causal effects, 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 $\mathbb{I}\{D_i > 0\}$ as the endogenous variable, since, as we show below, our instruments shift exposure to incarceration through both the extensive and intensive margins.

Technical probation violations and competing risks

Virtually all individuals *not* sentenced to incarceration are instead given a probation term that involves close supervision and often also restrictions on alcohol and drug use, work and socializing, and travel, as well as requiring payment of court fees and fines.[7] 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 explicity.

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.[8]

To overcome the dilemma of whether to include probation revocations in reoffending measures or not, we use two approaches. First and foremost, we present estimates of the effects of incarceration on reoffending with and without including probation revocations. Second, in Appendix III, we report estimates assuming that the risks of probation revocations and 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. This is because $\mathbb{E}\left[Y_{i,t}(d)|R_{i,t}\right] = \mathbb{E}\left[Y_{i,t}(d)\right]$,

[7]Before the end of 2011, most individuals sent to incarceration in North Carolina were not supervised after release since parole was eliminate by the SSA. They were only returned to incarceration if convicted of a new criminal offense.

[8]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".

III Setting and data

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 introduced during 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.[9]

The SSA established separate misdemeanor and felony "grids" that determine sentences as a function of offense severity and the offender's criminal history.[10] The SSA also eliminated parole by requiring that defendants serve the entirety of a minimum sentence. After doing so, defendants become eligible for early release, but can serve no more than 120% of their minimum sentence. Under the SSA, 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 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."[11] This is visually illustrated in Figure 2.1, 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;[12] and (iii) community

[9]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, Morenoff, Nguyen, and Bushway (2018) use similar designs to examine the effects of different criminal sanctions (e.g., prison vs. probation) on recidivism.

[10]Driving while impaired (DWI) offenses have separate sentencing grids.

[11]The maximum and minimum sentences are specified for three different ranges: Aggravated, presumptive, and mitigated. The majority of crimes are sentenced in the presumptive range.

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

punishment, or regular probation. These sentence types are denoted with "C/I/A" lettering at the top of each cell in the grid.

The combination of shifts in required sentence lengths and allowable sentence types generates large differences in punishments meted out across the grid. 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 also generates meaningful changes in the intensive margin as well. The grid has been modified occasionally since its introduction, which also generates variation in sentences. Specifically, we exploit a 2009 reform that substantially modified the mapping between prior record points and grid placement to validate our research design.

Data sources

We use administrative criminal justice records from two sources in North Carolina. 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 (civil infractions, most misdemeanors, etc.). We use this data 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 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.

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.[13]

Our research design utilizes discontinuities in sentencing guidelines. As a consequence, the analysis sample is restricted to individuals convicted for 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 classes of offense severity there are discontinuities both in the type and length of punishment, as is discussed in Section IV. However, when using Class D and C, there are only discontinuities in the guidelines with respect to the intensive margin, the length of punishment and no discontinuities in the extensive margin of the punishment type.[14] In addition, we also restrict the analysis to individuals aged between 15 and 65 at the time of offense.

Descriptive statistics

Summary statistics for our sample are presented in Table 2.1. On average, offenders are predominately male, roughly 50% black, and 30 years old (median age is 28) at the time they committed a felony offense. More than two-thirds of cases do not result in prison or jail sentences, and incarceration sentences average about 4.4 months. Conditional on receiving an incarceration sentence the average length is 13.6 months.

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.

---

[13]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 succeeds in estimating 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.

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

IV Empirical strategy

Within each SSA offense class (e.g., H, E, G) there are six prior record levels, each of which contains 4-5 prior points. Given five total jumps between prior record levels and five felony classes we include in the analysis, there are a total of 25 such discontinuities. In practice, we ignore the first discontinuity in each class, since there is only one prior point (for the pre-2009 grid, at least) in the prior record level to the left of the discontinuity. This leaves us with 20 total discontinuities. Within a class, each discontinuity can be thought of as a "mini" RD-style design.

Our setting is not a classic RD scenario with a continuous running variable, like a congressional election (Lee, 2008) or a college loan program (Solis, 2017), but rather a scenario with a discrete running variable that has 4 or 5 unique values before and after each discontinuity. The design exploits extreme non-linearities in sentencing within each felony class that are a function of prior points and are orthogonal to offenders' latent propensity to commit crime. Other studies that utilize non-linearities in assignment mechanisms include Kuziemko (2013) for the case of parole and Clark and Del Bono (2016) for school district allocation. Clark and Del Bono used non-linearities in the assignment formula to construct a "parameterized regression kink design." Analogously, each of our instruments can be thought of as providing a parametrized RD design.

Our preferred estimator "stacks" all the variation across each discontinuity in punishment to estimate a single treatment effect in a model that includes separate linear slopes in each cell of the sentencing grid. Treatment effect heterogeneity (and/or non-linearity) implies that the estimator identifies a weighted average of causal effects for different populations of compliers from the different discontinuities. Consider the two-equation system below. Equation (2.7) (first stage) estimates length of incarceration D_i as a function of prior points, convicted charge severity, punishment discontinuities, and other covariates. Equation (2.8) represents the relationship between reoffending within t months from conviction $Y_{i,t}$, incarceration and grid (and offender) controls. This system of equations is estimated using 2SLS.

$$D_i = \underbrace{\eta^1_{c_i} + X'_{it}\alpha_1}_{\text{Baseline controls}} + \underbrace{\sum_{k\in classes} 1\{class_i = k\}\left[\sum_{l\in thresh} \beta^2_{lk} 1\{p_i \geq l\}\,(p_i - l + 0.5) + \beta^1_k p_i\right]}_{\text{Linear slopes in prior points by class and level}} \tag{2.7}$$

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

$$Y_{i,t} = \beta_0 D_i + \underbrace{\eta^1_{c_i} + X'_{it}\alpha_1}_{\text{Baseline controls}} + \underbrace{\sum_{k \in classes} 1\{class_i = k\} \left[\sum_{l \in thresh} \beta^2_{lk} 1\{p_i \geq l\} (p_i - l + 0.5) + \beta^1_k p_i \right]}_{\text{Linear slopes in prior points by class and level}} \tag{2.8}$$

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

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 *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 discontinuity (the $1\{p_i \geq l\} (p_i - l + 0.5)$ effects), we recenter by $l - 0.5$ so that we measure the size of each discontinuity at the midpoint between the points as implied by the linear fits on either side, rather than at either extreme.[15] Figure 2.2 visually illustrates the above first stage specification, Equation (2.7), when focusing on class F.

The specification above stacks all the indicators for each discontinuity (that is, the indicators $\sum_{k \in classes} \sum_{l \in [5,9,15,19]} \gamma_{kl} 1\{c_i = k\} 1\{p_i \geq l\}$) 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 2.1. For the main analysis, we use these five punishment type discontinuities, which provide the most salient changes in sentences (red lines in Figure 2.1). When exploring heterogeneity in treatment effects, we also use the other 15, discontinuities although results are similar regardless of the instrument set used.

First stage effects of grid discontinuities

This research design captures large discontinuities in sanctions across the sentencing grid. For example, Figure 2.2 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 if shifted from 8 prior points to 9, which determines whether the offender is classified to prior record level III or IV. Note that this variation in punishment type falls at different points in the range of prior record points depending on the offense class. For example, in class H, which contains the most defendants in the data, unlike class F, the change in punishment type (Intermediate vs. Active) falls between prior record levels V and

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

VI, which generates an extensive margin discontinuity between prior record points 18 and 19.

Another way to visualize the shifts in time served due to our instruments is to plot the estimated weights of the average causal response from Equation (2.2). The estimated probabilities $\Pr(D_i(1) \geq d > D_i(0))$ capture the distributional shifts in exposure to incarceration caused by each of the binary instruments and can be estimated as $\mathbb{E}\left[1(D_i \geq d)|Z_i = 1\right] - \mathbb{E}\left[1(D_i \geq d)|Z_i = 0\right]$. Figure 2.2 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.

The estimates of the probabilities $\Pr(D_i(1) \geq d > D_i(0))$ also provide a test for the monotonicity assumption (Angrist and Imbens, 1995). If the instruments satisfy the monotonicity assumption then $\widehat{\Pr}(D_i(1) \geq d > D_i(0))$ should never cross the zero line, since a probability cannot have a negative value.

In all the regression specifications employed below, we control in a flexible way for the offender's criminal history using not only the flexible linear controls in prior points implicit in the RD specification, but also indicators for any previous incarceration spell, 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.

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 important 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 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 to all offenders.[16]

Figure 2.3 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 in magnitude and we cannot reject zero change in the risk score. A Wald test for the joint

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

significance of all five discontinuities also easily 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 2.2.

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.

V 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.

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 above relative to below the punishment type discontinuity on various outcomes. Figure 2.4 shows that individuals to the right of the discontinuity have a sharp drop in their likelihood of being reincarcerated within three years of conviction of roughly 11.9 p.p. 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 2.5 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. Individuals 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 2.5 also shows that the discontinuities stop predicting incarceration status primarily as a result of the initially incarcerated offenders being released and not because the offenders initially sentenced to probation are being incarcerated.[17]

Next we examine the dynamic effects of incarceration on reoffending and incapacitation across offenders from all felony classes (analysis by felony class is further discussed in

[17]The reduced form patterns documented in Figures 2.4 and 2.5 are similar across the other felony classes as is shown in Appendix Figures A.14, A.15, A.16, A.17, and A.18.

Appendix IV). We estimate Equation (2.7) while imposing that the coefficients on the indicators for being above a punishment type discontinuity are all equal, $\gamma^2_{E,4} = \gamma^2_{F,9} = \gamma^2_{G,14} = \gamma^2_{H,19} = \gamma^2_{I,9} = \gamma^{RF}$. This strategy averages across all five offense classes in our analysis dataset, but collapses our variation into a single coefficient. The parameter γ^{RF} can be thought of as the average reduced form effect across the five punishment discontinuities.[18]

Figure 2.6 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 the effect of being to the right of a punishment type discontinuity on outcomes *within* a single month from conviction. The estimates in Figure 2.5, therefore, represent four points from the blue line in Panel (a) (but only for class F offenders).

The discontinuities cause a large and immediate effect on 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. This confirms the findings from Figure 2.5, Panel (d).

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 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, which is not the case.

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 2.6. This graph shows 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 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 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. 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

[18] 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 similar results.

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 (see Appendix Figure A.9).

These estimates recover policy relevant parameters for optimal crime control policy. They estimate the 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.

In a first attempt to isolate behavioral responses from incapacitation effects, Panel (b) of Appendix Figure A.9 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 2.6, the discontinuities no longer influence the likelihood of being behind bars 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.[19]

2SLS estimates

We next present 2SLS estimates using length of incarceration as the endogenous regressor of interest. Table 2.2 contains results for committing any new offense within 3 years of conviction. Column 1 shows that the OLS estimate is negative and suggest that a one year incarceration spell reduces reoffending by 12.9%. 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. 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). The 2SLS 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.

To investigate the effects on different types of reoffending, in Table 2.3 we report 2SLS estimates for indicators of committing different types of offenses within three years of conviction. The effects of one year of incarceration are similar in sign and magnitude, relative to the overall mean, across the different types of reoffending. For example, a one

[19]The approach of estimating incarceration effects on reoffending using a measure of crime that includes only periods of time post-conviction in which the instrument stop being predictive of incarceration status was first proposed by Bhuller et al. (2018b), who study incarceration and recidivism in Norway.

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%).

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 2.3 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 2.2, illustrates the importance of carefully accounting for probation revocations. The larger effects for reincarceration are mostly due to probation revocations that do not result in a new offense but nevertheless lead to reincarceration. In Appendix III, we discuss several solutions to the problem of bias due to probation revocations that are 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 assumptions that show incarceration length has 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, these estimates non-parametrically identify a weighted average of local average treatment effects (i.e., the average causal response). However, it is also of interest to examine whether the simple model of a constant (no heterogeneity) 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 2.2 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.10 reports 2SLS estimates for reoffending within one year. It is clear that now that the J-statistics are significant for almost all the offense types and the null of a 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 3.

Treatment effect heterogeneity

Treatment effect heterogeneity by felony class is discussed in Appendix IV. Overall, the patterns in all the classes look similar, although there is substantial variation in the duration of time that the discontinuity has a meaningful effect on incarceration status, i.e., in the shift to incarceration length caused by the 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-release, there is some suggestive evidence that only longer sentences persistently reduce it.

Incapacitation effects and selection to incarceration

The magnitude of the incapacitation effects largely reflects the average risk of the population sentenced to incarceration. An interesting question is 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

$$\frac{\mathbb{E}\left[Y_{i,t}\cdot\left(1-1(D_i>0)\right)|Z_i=1\right]-\mathbb{E}\left[Y_{i,t}\cdot\left(1-1(D_i>0)\right)|Z_i=0\right]}{\mathbb{E}\left[1-1(D_i>0)|Z_i=1\right]-\mathbb{E}\left[1-1(D_i>0)|Z_i=0\right]}=\mathbb{E}\left[Y_{i,t}(0)|D_i(1)>0=D_i(0)\right] \tag{2.9}$$

Figure 2.7 shows estimates of $\mathbb{E}\left[Y_{i,t}(0)|D_i(1)>0=D_i(0)\right]$ (i.e., the compliers) and $\mathbb{E}\left[Y_{i,t}|D_i=0\right]$ (i.e., all those not incarcerated) 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.

Estimates from “at-risk”

In this section, we turn to an approach commonly used in the literature to separate the effect of incarceration exposure on reoffending after release from any incapacitation effect due to the initial sentence. This approach measures reoffending since each individual’s “at-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, the measure starts at the day of release from incarceration. Any differences in offending between the two groups using this at-risk 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 2.4 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 2.4. 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.

VI Threats to identification

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 would potentially bias our estimates. To address this potential concern, we compare our estimates of incarceration effects on reoffending using the offense class of each individual's most severe *charge* instead of the most severe *conviction* to define the instruments. Since the most severe charge is determined at arraignment, it is unlikely to be affected by plea negotiation. Appendix Table A.12

Tables and Figures

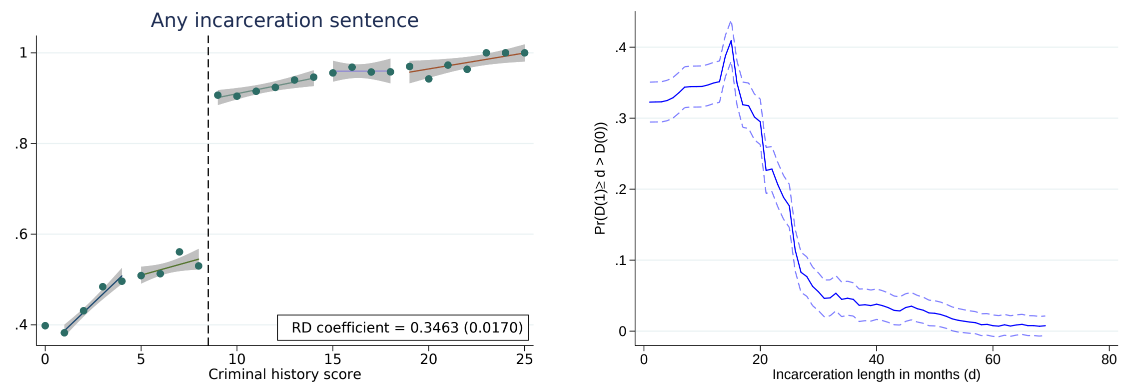
Figure 2.1: Sentencing guidelines until 2009 (illustration)

	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 official sentencing guidelines with illustrations for clarifications. The rows of the table refer to categories indicating the severity class of the felony offense that the individual was convicted of. The columns indicate the criminal history. Each individual is assigned a prior record points score that is a weighted average of past convictions based on severity. Based on the prior points score offenders are classified into prior record levels (columns) according to legislated thresholds. sets minimum sentences for each offense class and prior record level combination, which we refer to as a grid “cell." 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. 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". The red lines indicate critical places in the grid at which transitions across columns (prior record level) cause a change in the recommended sentence type. Indicators for these five transitions, marked with the red line, are our core instruments.

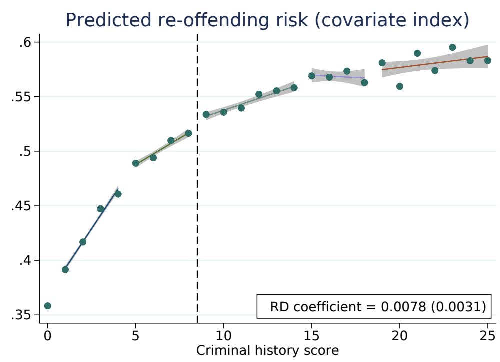
Figure 2.2: First stage: sentencing outcomes by prior points for offenders convicted of a class F felony offense



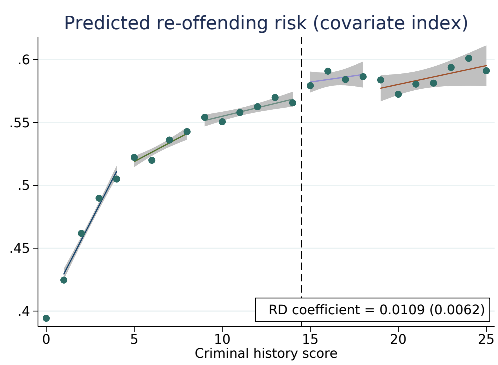
(a) Any incarceration (b) Distributional effects ($\Pr(D_i(1) \geq d > D_i(0))$)

Notes: This figure shows the first stage effect of the punishment type discontinuity in class F on probability to be incarcerated and on the length of incarceration. In Panel (a), the x-axis is the number of prior record points. The y-axis is the share of offenders who are sentenced to an active incarceration punishment. In Panel (b), the x-axis reports different incarceration lengths (d). The y-axis plots the estimate of the probability that an individual will serve less than d months if he is to the left of the discontinuity and atleast d months if he is to the right (i.e., $\Pr(D_i(1) \geq d > D_i(0))$).

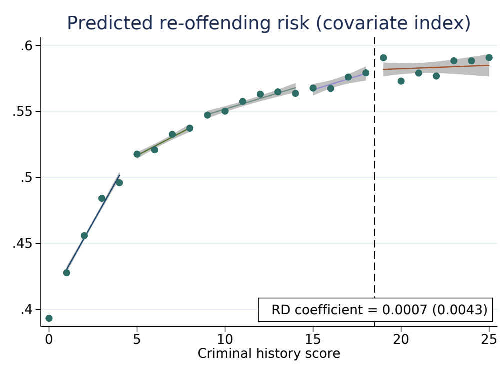
Figure 2.3: Predicted reoffending score by felony offense severity class and prior points



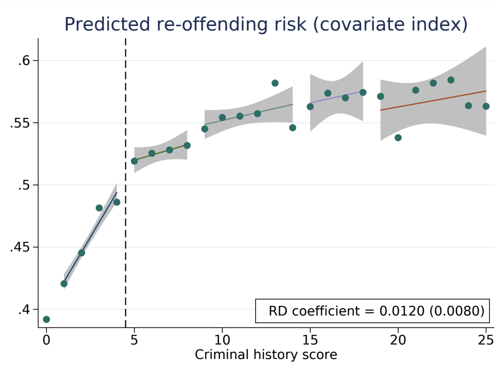
(a) Class I



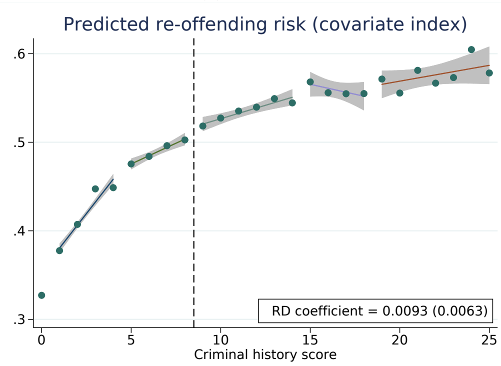
(b) Class G



(c) Class H



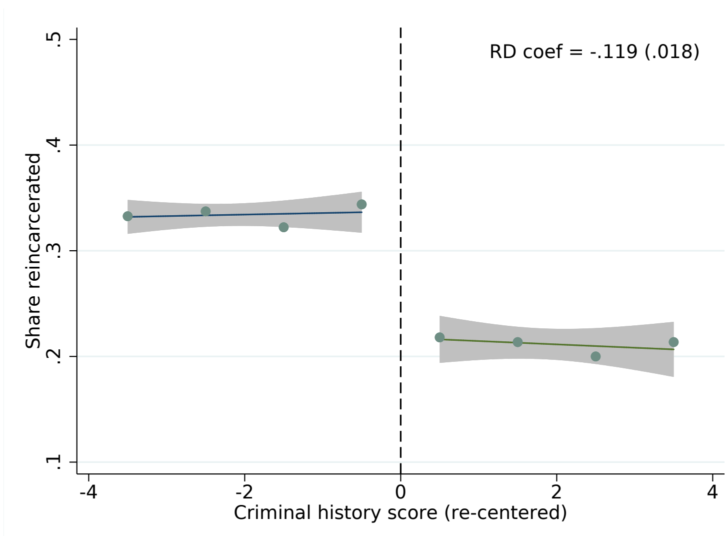
(d) Class E



(e) Class F

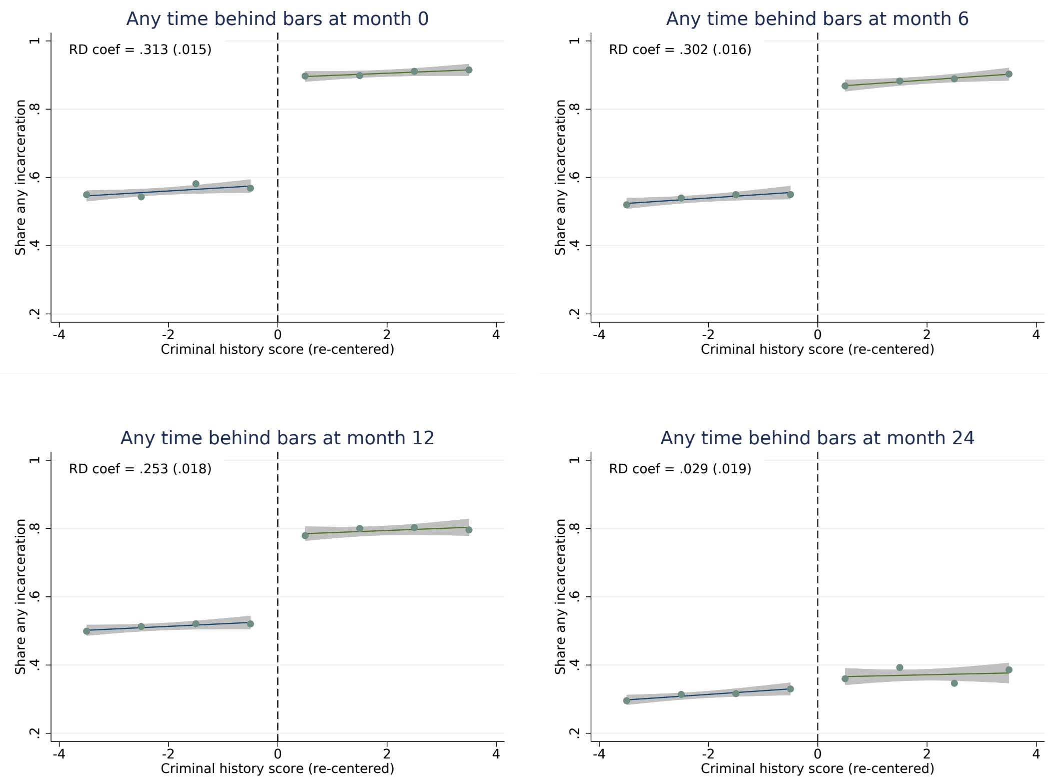
Notes: This figure shows how a summary index (i.e., score) of the covariates varies smoothly across the punishment type discontinuities in each offense class. The x-axis in all the plots is the number of prior record points. The y-axis shows the average predicted reoffending score that summarizes the predictions of all the covariates (e.g., age, race, criminal history) on reoffending within 3 years of the time of release to the community, i.e., 3 years from at-risk. 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; Card et al., 2015; Londono-Velez et al., 2018). Standard errors are clustered at the individual level.

Figure 2.4: Share reincarcerated within three years of conviction (class F felony offenses)



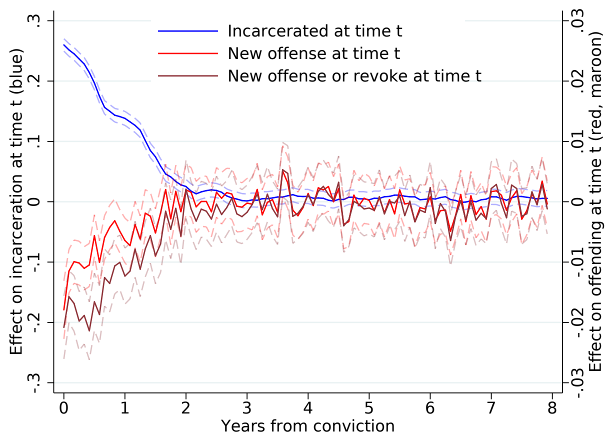
Notes: This figure shows reduced form estimates of being to the right of a punishment type discontinuity on the likelihood of reincarceration within three years from the date of conviction. The x-axis shows the recentered value of prior record points. The y-axis reports the share of individual who have been reincarcerated within three years of conviction. Our parameter of interest, which is reported in the figure (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 2.5: Dynamic differences in incarceration status *at* a given month after conviction (class F felony offenses)

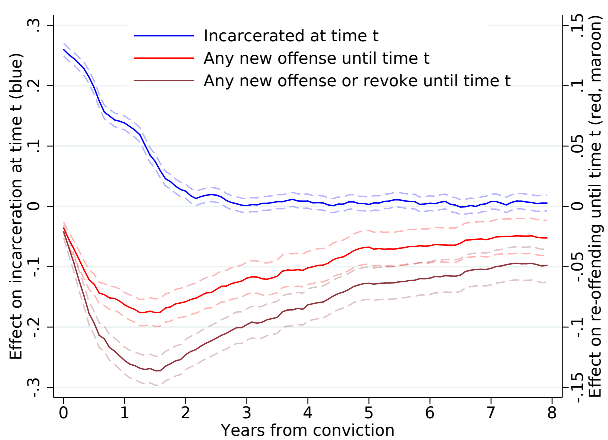


Notes: This figure shows reduced form RD estimates of being to the right of the punishment type discontinuity in class F on the likelihood of being incapacitated behind bars at month t after the date of conviction. The x-axis shows the recentered value of prior record points. The y-axis reports the share 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.

Figure 2.6: Effects on reoffending *at* period t from conviction and on *any* reoffending up to period t from conviction



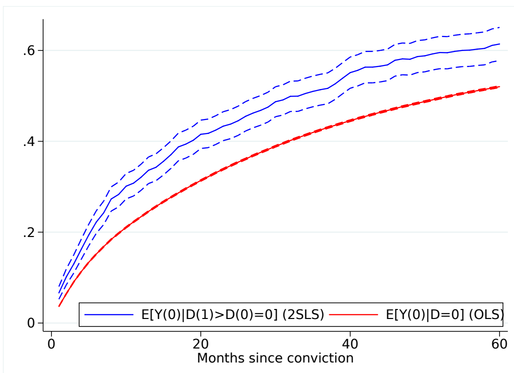
(a) Reoffending at period t from conviction



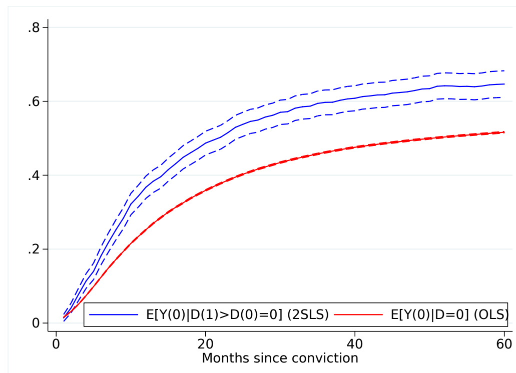
(b) Any reoffending until period t from conviction

Notes: This figure shows reduced form RD 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. For example, Figure 2.5 shows in detail four points on the blue line at 0, 6, 12, and 24 months from conviction when focusing only on individuals convicted of a class F felony offense. 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 ever committing *any* new offense *until* month t from conviction, and the maroon color line (right y-axis) the estimates when also including probation revocations as offending. Standard errors are clustered by individual.

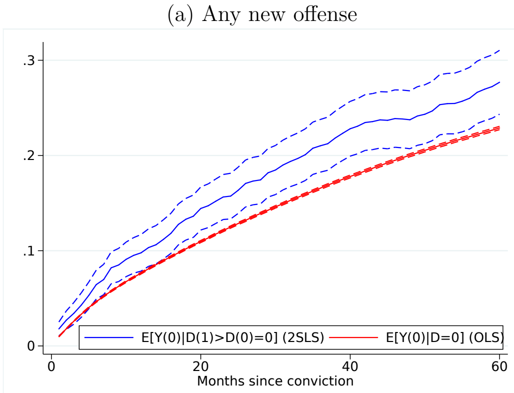
Figure 2.7: A comparison of reoffending patterns between non-incarcerated and compliers under the probation (non-incarceration) regime



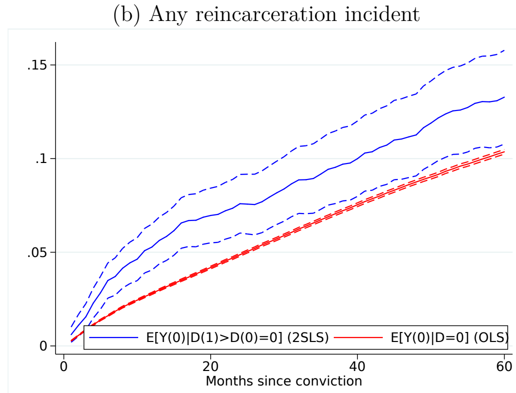
(a) Any new offense



(b) Any reincarceration incident



(c) Any new drug offense



(d) Any new assault offense

Notes: This figure shows the likelihood of reoffending among individuals who are not sentenced to incarceration. The OLS estimate (red line) is the reoffending probability of offenders who have not been incarcerated, i.e., those sentenced to probation ($\mathbb{E}\left[Y_{i,t}(0)|D_i = 0\right]$). The blue line reports estimates of the probability of reoffending for complier, i.e., those are offenders who are not incarcerated if they are to the left of a punishment type discontinuity but will be incarcerated if they are to the right ($\mathbb{E}\left[Y_{i,t}(0)|D_i(1) > D_i(0) = 0\right]$). 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).

Table 2.1: 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: The table shows summary statistics of the analysis sample. Notice that offenders tend to serve longer than their sentences.

Table 2.2: 2SLS estimates of effect of months of incarceration on committing any new offense within three years of sentencing

	(1) OLS	(2) OLS	(3) RD	(4) RD
Months incap	-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)	-0.129	-0.152	-0.203	-0.224
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: 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 (in parentheses) are clustered by individual. The OLS estimates in Columns 1 and 2 are from estimating Equation (2.8) using OLS. The 2SLS estimates in columns 3 and 4 are from estimating Equations (2.8) and (2.7) using 2SLS. The J stat refers to the Sargan-Hansen test of over-identifying restrictions. The test examines the null hypothesis that incarceration has the same effects when estimated using the different instruments under the assumption that the effects of incarceration 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.

Table 2.3: 2SLS estimates of effect of months of incarceration on various reoffending outcomes within three years of sentencing

	Measure of crime					
	(1)	(2)	(3)	(4)	(5)	(6)
	Re-incar	Any new offense	Felony	Assault	Property	Drug
Months incap	-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 for non.incar	0.462	0.425	0.306	0.0690	0.164	0.166
One year effect in percentages	-0.360	-0.224	-0.236	-0.376	-0.239	-0.203
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: 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 (in parentheses) are clustered by individual. Each column represents a different type of new offense (e.g., drug, property). For example, the estimates in Column 2 are the same as the estimates in Column 4 of Table 2.2. The estimates in each column are from Equations (2.8) and (2.7). The J stat refers to the Sargan-Hansen test of over-identifying restrictions. The test examines the null hypothesis that incarceration has the same effects when estimated using the different instruments under the assumption that the effects of incarceration 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. The Lochner-Moretti p-values are a generalization of the standard Hausman test of endogeneity to an ordered treatment with multiple levels.

Table 2.4: 2SLS estimates of effect of months of incarceration on various reoffending outcomes within three years of at-risk

	Measure of crime					
	(1)	(2)	(3)	(4)	(5)	(6)
	Re-incar	Any offense	Felony	Assault	Property	Drug
Months incap	-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 (non-incar)	0.462	0.425	0.305	0.0690	0.164	0.166
One year effect in %	-0.193	-0.0108	0.0222	-0.0557	0.0829	0.0857
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: 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 release date. Standard errors are clustered by individual. Each column represents a different type of new offense (e.g., drug, property).

Chapter 3

A Model of the Behavioral Effects of Incarceration

I A complete model of incarceration effects

Thus far, our analysis in Chapter 2 has investigated the reduced form effects of sentencing discontinuities 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 heterogeneity is present. Lastly, as is discussed in Section II (and Appendix II), 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 the 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. We estimate the model parameters via a semi-parametric two-step control function estimator for the effects of incarceration on the failure function of reoffending within t months from *at-risk*.

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. For example, job training programs typically involve a period where treated workers are either not working or working less due to the training, albeit the control workers are unconstrained, making it difficult to estimate the effect of the program on worker's wages or employment once training is complete (Ham and LaLonde, 1996; Eberwein et al., 1997). Many other settings—including other education treatments, unemployment spells, and military service—share a similar difficulty.

Selection to incarceration

We begin by describing the selection process into different spells of 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

$$D_i = d \quad \text{if} \quad \underbrace{C^l_{d-1}(Z^l_i)}_{\text{cut-offs}} \leq \underbrace{X_i'\gamma^l_0}_{\text{Observed component}} + \underbrace{\nu_i}_{\text{Unobserved component}} < C^l_d(\underbrace{Z^l_i}_{\text{Instrument}}) \tag{3.1}$$

where $\nu_i \sim N(0,1)$ and Z^l_i is an indicator for whether individual i is above or below 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

$$\begin{aligned} C^l_{d-1}(Z^l_i) \leq C_d(Z^l_i) \quad \forall Z^l_i,\ l \\ C^l_{-1}(Z^l_i) = -\infty, \quad C^l_{\bar{D}}(Z^l_i) = \infty \quad \forall Z^l_i,\ l \end{aligned} \tag{3.2}$$

The above model differs from a regular ordered probit model by allowing choice thresholds to depend on Z^l_i.[1] 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. This modeling therefore captures the variation that is introduced by the sentencing non-linearities in the grid.

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), however, we do not allow the thresholds to be stochastic since it does not seem necessary.

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

Reoffending and selection

We next model the likelihood of reoffending within t months from conviction and its relationship to the selection process for 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:

$$Y_{i,t}(d), \nu_i \perp\!\!\!\perp Z_i | X_i \quad \forall d \tag{3.3}$$

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 four 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 and unobserved factors ν_i, and (iv) time-varying controls $W_{i,d}$. The relationship between incarceration length and future criminality post release is captured by allowing the mean potential outcomes to depend on ν_i. We assume that for each level of initial incarceration d the conditional expectation of $Y_{i,t}(d)$ is[2]

$$\mathbb{E}\left[Y_{i,t}(d)|X_i, Z_i^l, W_{i,d}, \nu_i\right] = \underbrace{X_i'\xi_{t-d}}_{\text{Covariates pre-conviction}} + \underbrace{W_{i,d}'\eta_{t-d}}_{\text{Covariates at release}} + \underbrace{\overbrace{\theta^0_{d,t-d}}^{\text{Average effects}} + \overbrace{\theta^1_{d,t-d}\nu_i}^{\text{Heterogeneity}}}_{\text{Behavioral effects}} \tag{3.4}$$

where $t-d$ is the number of periods that the individual is at-risk to reoffend and $Y_{i,t}(d) = 0$ if $t-d \leq 0 \quad \forall i$, i.e., that during the period of time that the offender is removed from society she cannot commit crime.[3] Identification relies on the assumption that the model parameters are additively separable. This restriction is similar to assumptions in commonly used research designs such as the difference-in-differences approach with time-varying controls.[4]

Equation (3.4) 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 $t-d \leq 0 \quad \forall i$. 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, 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 to have a different effect on reoffending. Finally, heterogeneity with respect to the unobserved factors $\theta^1_{d,t-d}\nu_i$, captures differences in the effects of incarceration across individuals with

[2]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 effects the aging component by directly controlling for it.

[3]We follow the standard in the literature and focus on offenses that take place in non-institutionalized society (i.e., outside of prison) to define measures of reoffending.

[4]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.

varying levels of latent criminality as in Garen (1984) and Card (1999) who study the choice of years of education.

By iterated expectations, Equation (3.4) can be written as:

$$\mathbb{E}\left[Y_{i,t}|X_i, Z_i^l, W_{i,d}, D_i = d\right] = X_i'\xi_{t-d} + W_{i,d}'\eta_{t-d} + \underbrace{\theta^0_{d,t-d} + \theta^1_{d,t-d}\overbrace{\lambda_i\left(X_i, Z_i^l, d\right)}^{\text{Control function}}}_{\text{Behavioral effects}} \tag{3.5}$$

where $Y_{i,t}(d) = 0$ if $t - d \leq 0 \quad \forall i$ and $\lambda_i\left(X_i, Z_i^l, d\right) = \mathbb{E}\left[\nu_i|X_i, Z_i^l, D_i = d\right]$ is the generalized residual from the first stage, Equation (3.1). After fitting the first stage ordered choice model of time served, these generalized residuals are readily estimated. Equation (3.4) 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, Heckman, Meghir, and Vytlacil, 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 will also show that this simplified model still provides a good fit to the data and can also replicate the experimental variation produced by the instrumental variables.

Identification

Identification of $\theta^1_{d,t-d}$ relies on variation in Z_i^l given $X_i = x$, $W_{i,d} = w$, $D_i = d$, and t:

$$\begin{aligned}
&\mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,d} = w, D_i = d, Z_i^l = 1\right] - \mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,d} = w, D_i = d, Z_i^l = 0\right] \\
&= \theta^1_{d,t-d}\lambda(x,1,d) - \theta^1_{d,t-d}\lambda(x,0,d) \\
&\Rightarrow \theta^1_{d,t-d} = \frac{\mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,t} = w, D_i = d, Z_i^l = 1\right] - \mathbb{E}\left[Y_{i,t}|X_i = x, W_{i,t} = w, D_i = d, Z_i^l = 0\right]}{\lambda(x,1,d) - \lambda(x,0,d)}
\end{aligned} \tag{3.6}$$

The intuition behind the identification argument above is that given similar observables ($X_i = x$ and $W_{i,d} = w$) two individuals, one to the left and the other to the right of a discontinuity who both got the same incarceration spell must be different 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. Another interpretation of Equation (3.6) is as an infeasible 2SLS regression of $Y_{i,t}$ on $\lambda(\cdot)$ using Z_i^l 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_i^l. 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.

Model estimates

We first discuss estimates of the simplified version of Equation (3.5) 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.[5]

Table 3.1 shows the main estimates of the simplified version of Equation (3.5). 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 selection on unobservables shrinks the marginal effect calculated using OLS is half the size of the estimates using the control function. 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.

Behavioral effects for "average compliers" can be approximated by plugging in the model estimates the average value of ν for compliers at each of the discontinuities. Table 3.2 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 3.2 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 3.2, we report effects for "average compliers." Figure 3.1 shows the effects of different

[5] 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.

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.

Characterizing selection to incarceration

Our model can also be used to examine the selection process to incarceration. Whereas in other settings selection is informative about individuals' costs of take-up, in our context the assignment process describes the considerations that motivate judges. An important question, therefore, is whether judges sentence those with higher recidivism risk to longer incarceration spells (i.e., selection on levels), those likely to reduce reoffending the most because of prison (i.e., selection on gains), or both.

The relationship between incarceration length and unobserved criminality follows directly from the sign of the control function ($\hat{\nu}_i$) coefficient in Table 3.1. 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 the offenders who are more likely to reoffend.

We next examine whether there is evidence of selection on gains. Appendix Figure A.10 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.

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., $\mathbb{E}\left[Y_{i,t}|Z_i^l = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i^l = 0\right]$.[6]

Figure 3.2 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 (3.5). 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 the RD estimates well, $R^2 = 0.972$ and the slope coefficient is 0.87 with

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

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 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

$$\begin{aligned}\hat{\tau}(d) &\equiv \mathbb{E}\left[Y_{i,t}(d) - Y_{i,t}(d-1) | W_{i,d}, D_i(1) \geq d > D_i(0)\right] \\ &= \underbrace{\mathbb{E}\left[X_i' | D_i(1) \geq d > D_i(0)\right](\xi_{t-d} - \xi_{t-d+1}) + w'(\eta_{t-d} - \eta_{t-d+1})}_{\text{Effect of reduction in time at-risk}} \\ &\quad + \underbrace{(\theta^0_{d,t-d} - \theta^0_{d-1,t-d+1}) + (\theta^1_{d,t-d} - \theta^1_{d-1,t-d+1})\mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0)\right]}_{\text{Total behavioral effects}}\end{aligned} \tag{3.7}$$

The first term captures the reduction in reoffending expected from having one less month at risk 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 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

$$\begin{aligned}\hat{\tau}^{null}(d) &= \underbrace{\mathbb{E}\left[X_i' | D_i(1) \geq d > D_i(0)\right](\xi_{t-d} - \xi_{t-d+1}) + w'(\eta_{t-d} - \eta_{t-d+1})}_{\text{Effect of reduction in time at-risk (through covariates)}} \\ &\quad + \underbrace{(\theta^0_{0,t-d} - \theta^0_{0,t-d+1}) + (\theta^1_{0,t-d} - \theta^1_{0,t-d+1})\mathbb{E}\left[\nu_i | D_i(1) \geq d > D_i(0)\right]}_{\text{Effect of reduction in time at-risk under probation regime}}\end{aligned} \tag{3.8}$$

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 ACR weights for each discontinuity.

$$\text{Behavioral channel} = \sum_{d=1}^{\bar{D}} \left[\hat{\tau}(d) - \hat{\tau}^{null}(d)\right] \widehat{\Pr}\left(D_i(1) \geq d > D_i(0)\right) \tag{3.9}$$

The results of the above decomposition are presented graphically in Figure 3.3 for the simplified model described in Equation (3.5). The green line (square marker) represents the model replication of the reduced form effects (Equation 3.7), the black line (diamond marker) shows the model replication under the null of *no* behavioral effects (Equation 3.8), and the blue line (round marker) reports the estimates of the behavioral channel (Equation 3.9). 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 the different felony classes and the behavioral channel explains the majority of this reduction.

Table 3.3 summarizes the share of the effects explained by the behavioral channel. 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 back and re-interpret Figure 2.6. The control function estimates show that eight years after conviction the behavioral channel explains the majority of the observed crime-reducing effects.

Importantly, the above estimator of the behavioral effects in Equation (3.9) does not require extrapolating away from the individuals affected directly by the discontinuities, i.e., the compliers. The estimates in Table 3.1 report average treatment effects across all individuals; however, the above estimates from the decompositions show the behavioral effects for the populations of compliers that are directly influenced by the discontinuities in punishment at each felony class.

Probation revocations as non-random censoring

As discussed earlier, probation revocations can be viewed as a competing risk for reoffending. 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 reoffending 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 competing risk.

II Policy reforms

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.

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 does the marginal offense need 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 (2.8), 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

$$\text{Break-even value} \equiv \underbrace{\frac{\Delta\,(\text{Cumulative number of new offenses})}{\Delta\,(\text{Cumulative months incarcerated})}}_{\beta_{2SLS}} \times \frac{1}{\text{Incar. costs per month}} \tag{3.10}$$

Table 3.4 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 2.7 while imposing that being above a punishment discontinuity has the same effect across felony classes.[7]

[7]This can be thought of as a weighted average of the effects of the five punishment type discontinuities.

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. Nevertheless, the estimates suggest that the direct effects of incarceration on the marginal offenders' behavior are potentially insufficient to justify its use.

Extrapolating beyond the discontinuities

We begin by examining the optimality of the current sentencing guidelines and presenting suggestive evidence that there is potential for Pareto improvements. Our results 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 3.4 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, x-axis and 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 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.

III 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.[8] While our estimates show that incarceration sentences do not

[8]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/`

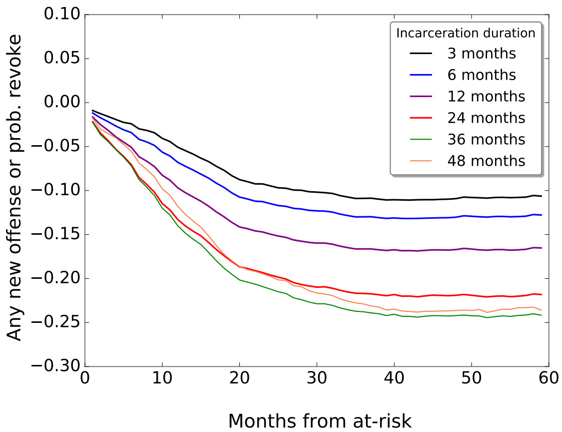
make offenders more likely to offend, they also demonstrate that incarceration has room to rehabilitate inmates further, especially when compared to carceral regimes in other developed countries. Since incarceration is unlikely to be abolished in the near future, understanding what features of imprisonment 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.

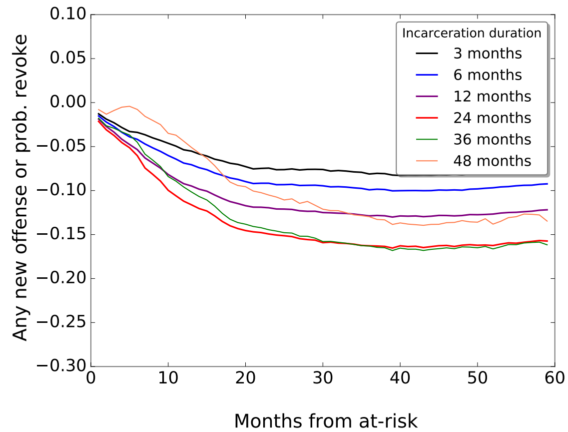
Tables and Figures

mandatory-minimum-drug-sentences-jeff-sessions-238295

Figure 3.1: Heterogeneity and non-linearity in the behavioral effects of incarceration on reoffending in the years after release



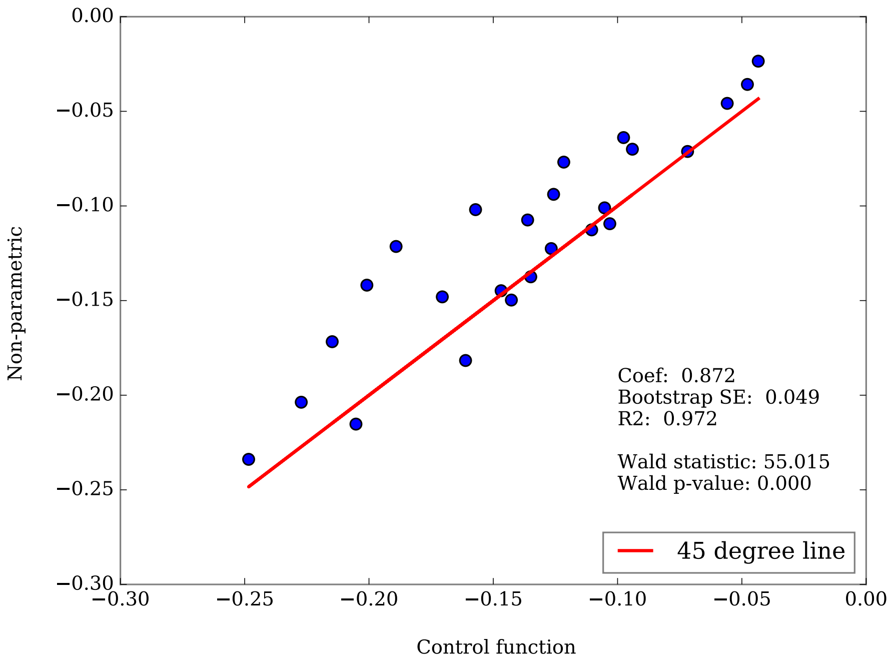
(a) Compliers of class I (least severe offenses)



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

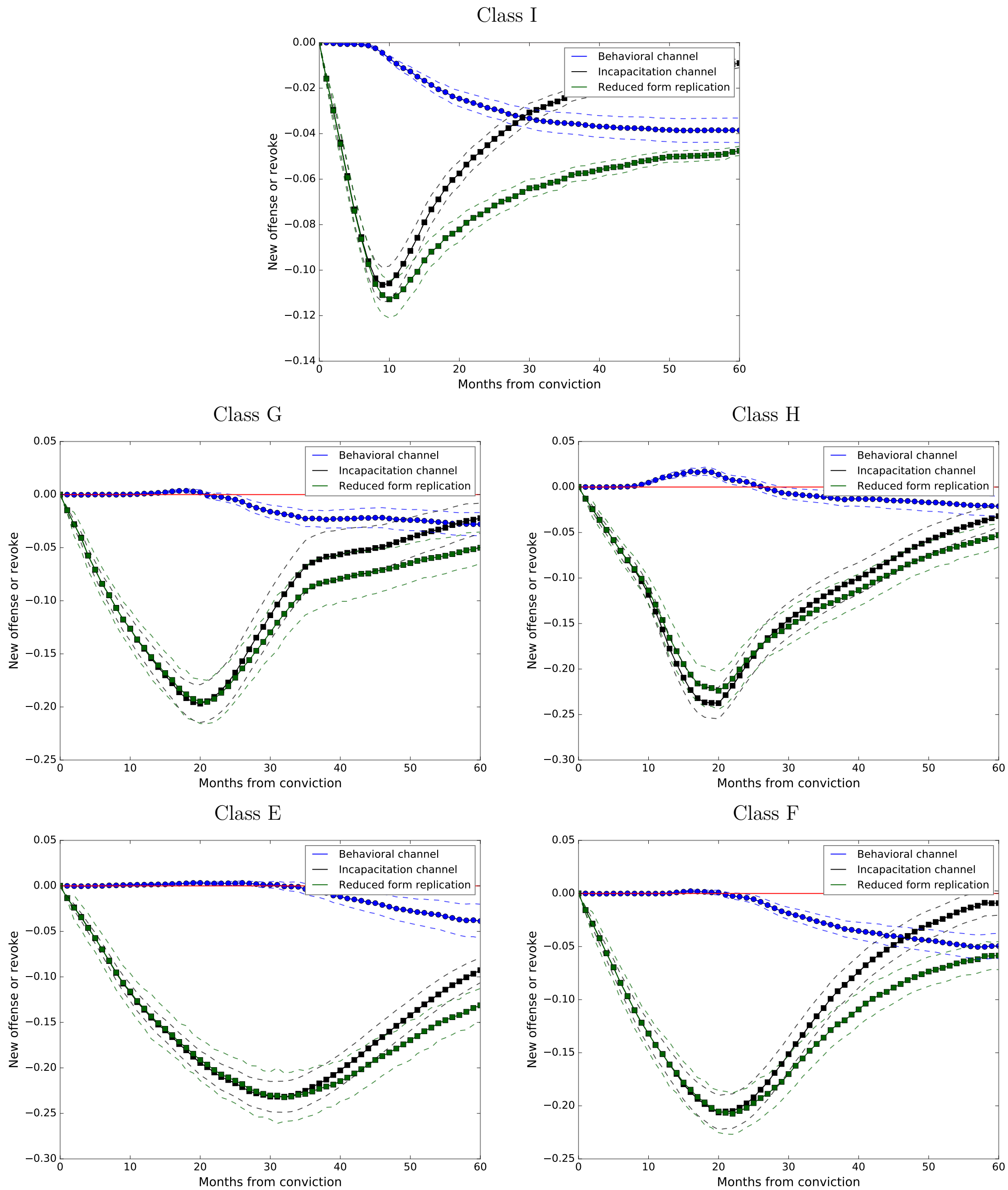
Notes: This figure shows the control function estimates of reoffending within t months from at-risk (i.e., release) for offenders with characteristics similar to those of compliers in class E (most severe offenses) and class I (least severe offenses). We use the average unobserved heterogeneity (ν) of 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 ν 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 II 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.

Figure 3.2: Control function goodness of fit: Replication of reduced form RD estimates of reoffending at various time windows from conviction



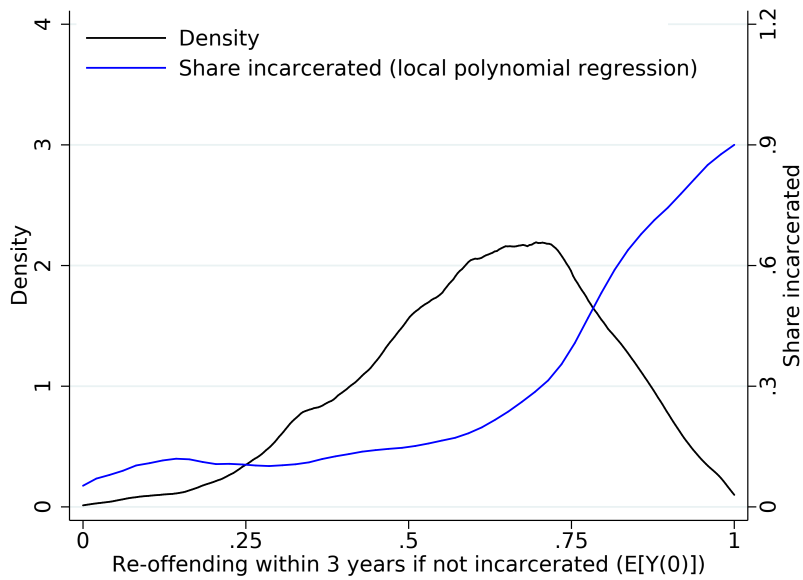
Notes: This figure shows how well the control function estimates of the model parameters, using the simplified specification that has a polynomial in D_i and an indicator for any incarceration sentence, can reproduce the quasi-experimental variation that is induced by the punishment type discontinuities in the sentencing guidelines (the red lines in Figure 2.1). The y-axis shows the non-parametric RD estimates of being to the right of a discontinuity relative to the left ($\mathbb{E}\left[Y_{i,t}|Z_i = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i = 0\right]$) on reoffending for each of the five felony classes for five time horizons (1,2,3,4, and 5 years from *conviction*) summing up to a total of 25 points. The x-axis represent the values of the replications of the reduced form estimates using the control function approach. The values represent the difference in reoffending (i.e., committing any new offense or probation revocation) within a given time horizon (e.g., 1 or 4 years) from conviction between individuals who are to the left of a discontinuity relative to those to the right. As was shown by Angrist and Imbens (1995) the reduced form of a treatment with multiple levels can be expressed as $\mathbb{E}\left[Y_{i,t}|Z_i = 1\right] - \mathbb{E}\left[Y_{i,t}|Z_i = 0\right] = \sum_{d=1}^{\bar{D}} \mathbb{E}\left[Y_{i,t}(d) - Y_{i,t}(d-1)|D_i(1) \geq d > D_i(0)\right] \Pr(D_i(1) \geq d > D_i(0))$. The red line shows the 45 degree line. If the control function approach perfectly replicates the reduced form RD estimates then all the points should be 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-square=1). A comparison of reduced form RD estimates from sentencing to the model based reconstructed reduced-form estimates jointly tests the goodness of fit of the selection model described by the ordered-choice model and the parametric restrictions imposed on $\mathbb{E}\left[Y_{i,t}(d)\right]$ 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 2SLS do not include adjustment for time-varying factors as is discussed in the Section II.

Figure 3.3: Decomposition by offense class of reduced form RD estimates into 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.

Figure 3.4: 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 for each level of predicted risk of reoffending. The x-axis show the predicted likelihoods 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 who have been incarcerated as a function of the reoffending likelihood (x-axis).

Table 3.1: Control function estimates of behavioral effects: 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 incap	-0.0611***	-0.0128	-0.0705***	-0.0832***	-0.0837***
	(0.00400)	(0.175)	(0.00427)	(0.00684)	(0.00686)
Years incap 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 $\times\, \hat{\nu}$				-0.0191**	-0.0229**
				(0.00640)	(0.00765)
Years incap $\times\, \hat{\nu}$				-0.0121*	-0.0125*
				(0.00511)	(0.00515)
Years incap square $\times\, \hat{\nu}$				-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. Standard errors are clustred by individual. The estimated control function $\widehat{\lambda}_i\left(X_i, Z_i^l, d\right) = \mathbb{E}\left[\nu_i | X_i, Z_i^l, D_i = d\right]$ 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 is from a 2SLS of each of the above control function specifications when the endogenous variables are only the $\hat{\nu}$ terms. Since our identifiying assumption imply that given ν the other variables in the model can be treated as "exogenous".

Table 3.2: Heterogeneity in model estimates of marginal behavioral effects: Any new offense or probation revocation within 3 years of at-risk

	Population	Compliers				
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Class E	Class F	Class G	Class H	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)
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. All the estimates are based on the specification in Column 4 of Table 3.1. All the marginal treatment effects are expressed in % terms relative to the mean reoffending rate among non-incarcerated offenders. The outcome of interest is any new offense or a probation revocation within 3 years of at-risk. Column 1 reports population treatment effect, i.e., when $\nu = 0$. Columns 2-5 report estimates using the average unobserved heterogeneity (ν) 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 ν 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 II 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 3.3: Share of reduced form RD estimates attributable to behavioral channel

	Class E	Class F	Class G	Class H	Class I
	(1)	(2)	(3)	(4)	(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 to behavioral and incapacitation channels. Each cell shows the share of the reduced form estimates that is explained by the behavioral channel. The outcome in all the estimates is any new offense or probation revocation within 1,3 or 5 years from the date of conviction. The rows indicate whether the reoffending occurred within 1, 3 or 5 years from convictions, i.e., the number of years from conviction in which any reoffending is measured. The standard errors are calculated using a block bootstrap procedure, at the individual level, with 500 iterations.

Table 3.4: 2SLS break-even approach estimates: The dollar values of crime to society that are 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 break-even dollar values that society needs to assign to crime averted to justify the 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 offense (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 \$2738.1.

Bibliography

Alberto Abadie. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, pages 284–293, 2002.

Atila Abdulkadiroglu, Parag A. Pathak, and Christopher R. Walters. Free to choose: Can school choice reduce student achievement? Working Paper 21839, National Bureau of Economic Research, December 2015. URL `http://www.nber.org/papers/w21839`.

David Abrams and Ryan Fackler. To plea or not to plea: Evidence from north carolina. July 2017.

David Abrams and Albert Yoon. The luck of the draw: Using random case assignment to investigate attorney ability. *The University of Chicago Law Review*, 74(4):1145–1177, 2007.

David Abrams, Marianne Bertrand, and Sendhil Mullainathan. Do judges vary in their treatment of race? *The Journal of Legal Studies*, 41(2):347–383, 2012. ISSN 00472530, 15375366.

Jeff Adachi. Using checklists to improve case outcomes. *Champion*, 2015.

Amanda Agan and Sonja Starr. Ban the box, criminal records, and racial discrimination: A field experiment. *The Quarterly Journal of Economics*, 133(1):191–235, 2018. doi: 10.1093/qje/qjx028. URL `http://dx.doi.org/10.1093/qje/qjx028`.

Amanda Agan, Matthew Freedman, and Emily Owens. Is your lawyer a lemon? incentives and selection in the public provision of criminal defense. Working Paper 24579, National Bureau of Economic Research, May 2018. URL `http://www.nber.org/papers/w24579`.

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

Anna Aizer and Joseph J. Doyle, Jr. Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2):759–803, 2015b. doi: 10.1093/qje/qjv003. URL `+http://dx.doi.org/10.1093/qje/qjv003`.

Junius L. Allison. Relationship between the office of public defender and the assigned counsel system. *Valparaiso University Law Review*, 10(3):399–422, November 1976.

Joseph Altonji, Todd Elder, and Christopher Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184, 2005.

James Anderson and Paul Heaton. How much difference does the lawyer make? the effect of defense counsel on murder case outcomes. *Yale Law Review*, 122(1), 2012.

James M. Anderson, Jeffrey R. Kling, and Kate Stith. Measuring interjudge sentencing disparity: Before and after the federal sentencing guidelines. *The Journal of Law and Economics*, 42(S1):271–308, 1999. doi: 10.1086/467426.

Michael Anderson and Jeremy Magruder. Split-sample strategies for avoiding false discoveries. 2017.

Abhay Aneja and Carlos Fernando Avenancio-León. Credit-driven crime cycles: The connection between incarceration and access to credit. 2018.

Joshua D. Angrist 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*, 90(430):431–442, 1995. doi: 10.1080/01621459.1995.10476535. URL `https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476535`.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434): 444–455, 1996.

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

Carolina Arteaga. The cost of bad parents: Evidence from the effects of incarceration on children's education. Working paper, 2018.

Itai Ater, Yehonatan Givati, and Oren Rigbi. The economics of rights: Does the right to counsel increase crime? *American Economic Journal: Economic Policy*, 9(2):1–27, May 2017. doi: 10.1257/pol.20160027. URL `http://www.aeaweb.org/articles?id=10.1257/pol.20160027`.

Susan Athey, Dean Eckles, and Guido W. Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

James Austin 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.

Alessandro Barbarino and Giovanni Mastrobuoni. The incapacitation effect of incarceration: Evidence from several italian collective pardons. *American Economic Journal: Economic Policy*, 6(1):1–37, 2014. ISSN 19457731, 1945774X. URL `http://www.jstor.org/stable/43189364`.

Patrick Bayer, Randi Hjalmarsson, and David Pozen. Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics*, 124(1):105–147, 2009. ISSN 00335533, 15314650. URL `http://www.jstor.org/stable/40506225`.

Gary S. Becker. Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217, 1968.

Manudeep Bhuller, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad. Intergenerational effects of incarceration. *AEA Papers and Proceedings*, 108:234–40, 2018a. doi: 10.1257/pandp.20181005. URL `http://www.aeaweb.org/articles?id=10.1257/pandp.20181005`.

Manudeep Bhuller, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad. Incarceration, recidivism and employment. Working Paper 22648, National Bureau of Economic Research, 2018b.

Jake. Bowers and Ben B. Hansen. Attributing effects to a cluster randomized get-out-the-vote campaign. *Journal of the American Statistical Association*, 104(487):873–885, 2009.

Paolo Buonanno and Steven Raphael. Incarceration and incapacitation: Evidence from the 2006 italian collective pardon. *The American Economic Review*, 103(6):2437–2465, 2013. ISSN 00028282. URL `http://www.jstor.org/stable/42920656`.

Stephen V. Cameron and James J. Heckman. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106(2):262–333, 1998. ISSN 00223808, 1537534X. URL `http://www.jstor.org/stable/10.1086/250010`.

David Card. The causal effect of education on earnings. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3, Part A, chapter 30, pages 1801–1863. Elsevier, 1 edition, 1999.

David Card and Stefano DellaVigna. Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1):144–161, 2013. ISSN 00220515. URL `http://www.jstor.org/stable/23644706`.

David Card, David S. Lee, Zhuan Pei, and Andrea Weber. Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483, 2015. ISSN 1468-0262.

Pedro Carneiro, 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*, 44(2):361–422, 2003. doi: 10.1111/1468-2354.t01-1-00074. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2354.t01-1-00074`.

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

Cheng Cheng and Wei Long. Can the private sector provide better police services? 2017.

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

Alma Cohen and Crystal S. Yang. Judicial politics and sentencing decisions. *accepted, American Economic Journal: Economic Policy*, 2018.

Will Dobbie, 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 2018a. URL `http://www.aeaweb.org/articles?id=10.1257/aer.20161503`.

Will Dobbie, Hans Grönqvist, Susan Niknami, Mårten Palme, and Mikael Priks. The intergenerational effects of parental incarceration. Working Paper 24186, National Bureau of Economic Research, January 2018b. URL `http://www.nber.org/papers/w24186`.

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

Curtis Eberwein, 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*, 64(4):655–682, 1997. ISSN 00346527, 1467937X. URL `http://www.jstor.org/stable/2971734`.

Sarah M. Estelle and David C. Phillips. Smart sentencing guidelines: The effect of marginal policy changes on recidivism. *Journal of Public Economics*, 164:270 – 293, 2018. ISSN 0047-2727.

Benjamin Feigenberg and Conrad Miller. Racial divisions and criminal justice: Evidence from southern state courts. Technical report, National Bureau of Economic Research, 2018.

Erica Field. Educational debt burden and career choice: Evidence from a financial aid experiment at nyu law school. *American Economic Journal: Applied Economics*, 1(1): 1–21, 2009. ISSN 19457782, 19457790. URL `http://www.jstor.org/stable/25760145`.

Joshua B. Fischman. Estimating preferences of circuit judges: A model of consensus voting. *The Journal of Law and Economics*, 54(4):781–809, 2011. ISSN 00222186, 15375285. URL `http://www.jstor.org/stable/10.1086/661512`.

Jean-Pierre Florens, 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*, 76(5):1191–1206, 2008.

Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Decomposition methods in economics. volume 4A, chapter 01, pages 1–102. Elsevier, 1 edition, 2011. URL `https://EconPapers.repec.org/RePEc:eee:labchp:4-01`.

Catalina Franco, David J. Harding, Jeffrey Morenoff, and Shawn D. Bushway. Estimating the effect of imprisonment on recidivism: Evidence from a regression discontinuity design. Accessed 9/30/2018 from https://catalinafranco.com/research/, 2017.

Peter N. Ganong. Criminal rehabilitation, incapacitation, and aging. *American Law and Economics Review*, 14(2):391–424, 2012. doi: 10.1093/aler/ahs010. URL `http://dx.doi.org/10.1093/aler/ahs010`.

John Garen. The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52(5):1199–1218, 1984. URL `https://EconPapers.repec.org/RePEc:ecm:emetrp:v:52:y:1984:i:5:p:1199-1218`.

Alexander Gelber, Adam Isen, , and Jae Song. The effect of pension income on elderly earnings: Evidence from social security and full population. 2016.

Christian Gourieroux, Alain Monfort, Eric Renault, and Alain Trognon. Generalised residuals. *Journal of Econometrics*, 34(1):5 – 32, 1987. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(87)90065-0. URL `http://www.sciencedirect.com/science/article/pii/0304407687900650`.

Donald P. Green and Daniel Winik. Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders. *Criminology*, 48(2): 357–387, 2010. ISSN 1745-9125. doi: 10.1111/j.1745-9125.2010.00189.x. URL `http://dx.doi.org/10.1111/j.1745-9125.2010.00189.x`.

William Greene and David Hensher. *Modeling Ordered Choices*. Cambridge University Press, 2010. URL `https://EconPapers.repec.org/RePEc:cup:cbooks:9780521194204`.

Jeffrey Grogger. The effect of arrests on the employment and earnings of young men. *The Quarterly Journal of Economics*, 110(1):51–71, 1995. URL `https://EconPapers.repec.org/RePEc:oup:qjecon:v:110:y:1995:i:1:p:51-71`.

John Ham and Robert LaLonde. The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica*, 64(1):175–205, 1996. URL `https://EconPapers.repec.org/RePEc:ecm:emetrp:v:64:y:1996:i:1:p:175-205`.

David J. Harding, 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*, 114(42):11103–11108, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1701544114. URL `http://www.pnas.org/content/114/42/11103`.

David J. Harding, Jeffrey D. Morenoff, Anh P. Nguyen, and Shawn D. Bushway. Imprisonment and labor market outcomes: Evidence from a natural experiment. *American Journal of Sociology*, 124(1):49–110, 2018. doi: 10.1086/697507. URL `https://doi.org/10.1086/697507`.

Caroline Wolf Harlow. *Defense counsel in criminal cases.* US Department of Justice, Office of Justice Programs, Bureau of Justice ..., 2001.

Oliver Hart, Andrei Shleifer, and Robert W. Vishny. The proper scope of government: Theory and an application to prisons. *The Quarterly Journal of Economics*, 112(4):1127–1161, 1997. URL `http://www.jstor.org/stable/2951268`.

John J. Haugh. The federal criminal justice act of 1964: Catalyst in the continuing formulation of the rights of the criminal defendant. *Notre Dame Law Review*, 41(6), 1966.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1912352`.

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

James J. Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738, 2005. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/3598865`.

James J. Heckman and Edward J. Vytlacil. Chapter 70 econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. volume 6, Part B of *Handbook of Econometrics*, pages 4779 – 4874. Elsevier, 2007a. doi: http://dx.doi.org/10.1016/S1573-4412(07)06070-9. URL `http://www.sciencedirect.com/science/article/pii/S1573441207060709`.

James J. Heckman and Edward J. Vytlacil. Chapter 71 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. volume 6, Part B of *Handbook of Econometrics*, pages 4875 – 5143. Elsevier, 2007b. doi: http://dx.doi.org/10.1016/S1573-4412(07)06071-0. URL `http://www.sciencedirect.com/science/article/pii/S1573441207060710`.

Randi Hjalmarsson. Juvenile jails: A path to the straight and narrow or to hardened criminality? *Journal of Law and Economics*, 52(4):779–809, 2009. URL `http://www.jstor.org/stable/10.1086/596039`.

Guido Imbens and Donald Rubin. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4):555–574, 1997.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

Radha Iyengar. An analysis of the performance of federal indigent defense counsel. Working Paper 13187, National Bureau of Economic Research, June 2007. URL `http://www.nber.org/papers/w13187`.

Judicial-Conference. Report of the judicial conference of the united states. Technical report, Administrative Office of the United States Courts, September 1952.

Luke Keele and Luke Miratrix. Randomization inference for outcomes with clumping at zero. *The American Statistician*, 0(ja):0–0, 2017. doi: 10.1080/00031305.2017.1385535. URL `https://doi.org/10.1080/00031305.2017.1385535`.

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133(1):237–293, 08 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx032. URL `https://dx.doi.org/10.1093/qje/qjx032`.

Patrick Kline and Christopher Walters. On heckits, late, and numerical equivalence. 2018.

Patrick Kline and Christopher R. Walters. Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, 131(4):1795–1848, 2016. doi: 10.1093/qje/qjw027. URL `http://dx.doi.org/10.1093/qje/qjw027`.

Jeffrey R. Kling. Incarceration length, employment, and earnings. *American Economic Review*, 96(3):863–876, 2006. doi: 10.1257/aer.96.3.863. URL `http://www.aeaweb.org/articles.php?doi=10.1257/aer.96.3.86`.

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

T. Kyckelhahn. Justice expenditures and employment, fy 1982-2007 statistical tables. Report NCJ 236218, U.S. Department of Justice, 2011.

Finbarr P. Leacy and Elizabeth A. Stuart. On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine*, 33(20):3488–3508, 2014.

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

Steven D. Levitt. The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. *The Quarterly Journal of Economics*, 111(2):319–351, 1996. doi: 10.2307/2946681. URL `http://qje.oxfordjournals.org/content/111/2/319.abstract`.

Katrine V. Løken, Magne Mogstad, and Matthew Wiswall. What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics*, 4(2):1–35, April 2012. doi: 10.1257/app.4.2.1. URL `http://www.aeaweb.org/articles?id=10.1257/app.4.2.1`.

Lance Lochner and Enrico Moretti. Estimating and testing models with many treatment levels and limited instruments. *The Review of Economics and Statistics*, 97(2):387–397, 2015. doi: 10.1162/REST_a_00475. URL `https://doi.org/10.1162/REST_a_00475`.

Charls E. Loeffer. Does imprisonment alter life course? evidence on crime and employment from a natural experiment. *Criminology*, 51(1):137–166, 2 2013. ISSN 1745-9125. doi: 10.1111/1745-9125.12000. URL `http:https://doi.org/10.1111/1745-9125.12000`.

Magnus Lofstrom and Steven Raphael. Crime, the criminal justice system, and socioeconomic inequality. *The Journal of Economic Perspectives*, 30(2):103–126, 2016. ISSN 08953309. URL `http://www.jstor.org/stable/43783709`.

Juliana Londono-Velez, 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.

Gary T. Lowenthal. Joint representation in criminal cases: A critical appraisal. *Virginia Law Review*, 64(7):939–989, 1978. ISSN 00426601. URL `http://www.jstor.org/stable/1072484`.

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

Eric Maurin and Aurelie Ouss. Sentence reductions and recidivism: Lessons from the bastille day quasi experiment. IZA DP No. 3990, February 2009.

Justin McCrary and Sarath Sanga. General equilibrium effects of prison on crime: Evidence from international comparisons. *Cato Papers on Public Policy*, 2, 2012.

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

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

Thomas J. Miles and Jens Ludwig. The silence of the lambdas: Deterring incapacitation research. *Journal of Quantitative Criminology*, 23(4):287–301, 2007. ISSN 07484518, 15737799. URL `http://www.jstor.org/stable/41954250`.

Stephen C. Moore. Conflicts of interest in public defender offices. *Journal of the Legal Profession*, 1984.

Michael Mueller-Smith. The criminal and labor market impacts of incarceration. Working paper, 2015.

Anita Mukherjee. Impacts of private prison contracting on inmate time served and recidivism. 2017.

David B. Mustard. Racial, ethnic, and gender disparities in sentencing: Evidence from the u.s. federal courts. *The Journal of Law and Economics*, 44(1):285–314, 2001.

Daniel S. Nagin and G. Matthew Snodgrass. The effect of incarceration on re-offending: Evidence from a natural experiment in pennsylvania. *Journal of Quantitative Criminology*, 29(4):601–642, 2013. ISSN 07484518, 15737799. URL `http://www.jstor.org/stable/43552154`.

National Center for State Courts. State sentencing guidelines profiles and continuum. Technical report, 2008.

Samuel Norris. Judicial errors: Evidence from refugee appeals. 2018.

Samuel Norris, Matthew Pecenco, and Jeffrey Weave. The intergenerational and sibling effects of incarceration: Evidence from ohio. 2018.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973.

Emily G. Owens. More time, less crime? estimating the incapacitative effects of sentence enhancements. *Journal of Law and Economics*, pages 551–579, 2009.

Jörg Paetzold and Hannes Winner. Bunching in the presence of deduction possibilities: Earnings and deduction responses to a large kink. Technical report, 2016.

Edward Prado. Process and progress: Reviewing the criminal justice act. *Law and Contemporary Problems*, 58(1), 1995.

Edward Prado, Robert Altman, Vincent Aprile, Judy Clark, Michael Davis, Edward Dennis, Robinson Everett, Thomas Hillier, George Revercomb, and Charles Ogletrel. Report of the committee to review the criminal justice act. Technical report, United States Judicial Conference, 1993.

Steven Raphael. *Controlling Crime: Strategies and Tradeoffs*, chapter Improving Employment Prospects for Former Prison Inmates: Challenges and Policy. University of Chicago Press, 2011.

Steven Raphael and Magnus Lofstrom. Incarceration and crime: Evidence from california's realignment sentencing reform. 2015.

Michael A. Roach. Indigent defense counsel, attorney quality, and defendant outcomes. *American Law and Economics Review*, 2014.

David Roodman. Impact of incarceration on crime. Technical report, 2017.

Paul R. Rosenbaum. The consquences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, 147(5):656–666, 1984.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2): 135–146, 1951.

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

Raphael Steven. The new scarlet letter? negotiating the u.s. labor market with a criminal record, 2014.

Megan Stevenson. Distortion of justice: How the inability to pay bail affects case outcomes. Working paper, 2016.

Megan Stevenson. Breaking bad: Mechanisms of social influence and the path to criminality in juvenile jails. *The Review of Economics and Statistics*, 99(5):824–838, 2017. URL `https://doi.org/10.1162/REST_a_00685`.

U.S. Department of Justice. National assessment of structured sentencing. Report, Bureau of Justice Assistance, 1996. URL `https://www.ncjrs.gov/app/publications/abstract.aspx?ID=153853`.

Edward Vytlacil. Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results. *The Review of Economics and Statistics*, 88(3):578–581, 2006. doi: 10.1162/rest.88.3.578. URL `https://doi.org/10.1162/rest.88.3.578`.

Jeffrey M. Wooldridge. Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445, 2015.

Crystal Yang. Free at last? judicial discretion and racial disparities in federal sentencing. *The Journal of Legal Studies*, 44(1):75–111, 2015. doi: 10.1086/680989. URL `https://doi.org/10.1086/680989`.

Crystal S. Yang. Local labor markets and criminal recidivism. *Journal of Public Economics*, 147:16 – 29, 2017. ISSN 0047-2727. doi: https://doi.org/10.1016/j.jpubeco.2016.12.003. URL `http://www.sciencedirect.com/science/article/pii/S0047272716302067`.

Mariyana Zapryanova. The efects of time in prison and time on parole on recidivism. 2017.

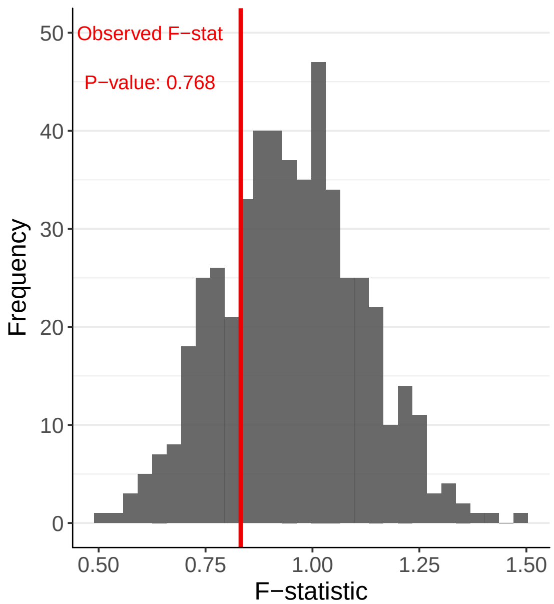
Appendix A

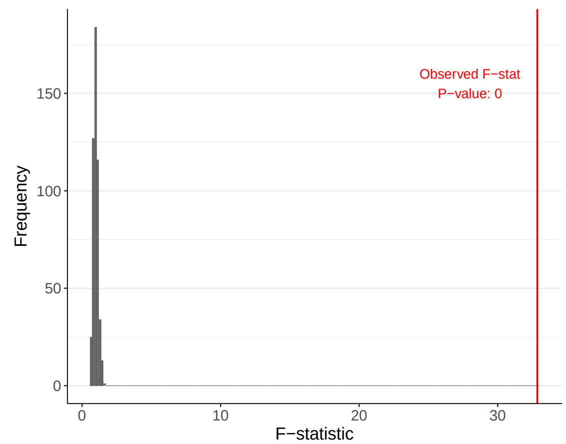
Chapter 1 appendices

Table A.1: Variation in defendant characteristics within a multiple defendant case

	Multiple	Co-defendants
Obs.	2.143	2.000
Black & Non-Black	0.222	0.214
Hispanic & Non-Hispanic	0.333	0.314
White & Non-White	0.245	0.236
Black & White	0.223	0.217
Black & Hispanic	0.073	0.067
White & Hispanic	0.081	0.073
Felony & Non-Felony	0.014	0.012
Prior arrest & No prior arrest	0.338	0.322
Prior conviction & No prior conv.	0.156	0.158
Prior incarceration & No prior incar.	0.287	0.280

Figure A.1: San Francisco: Monte-Carlo permutations of attorney type assignment within a case: F-statistic using Offense codes and controls





Within multiple defendant case Across single defendant cases

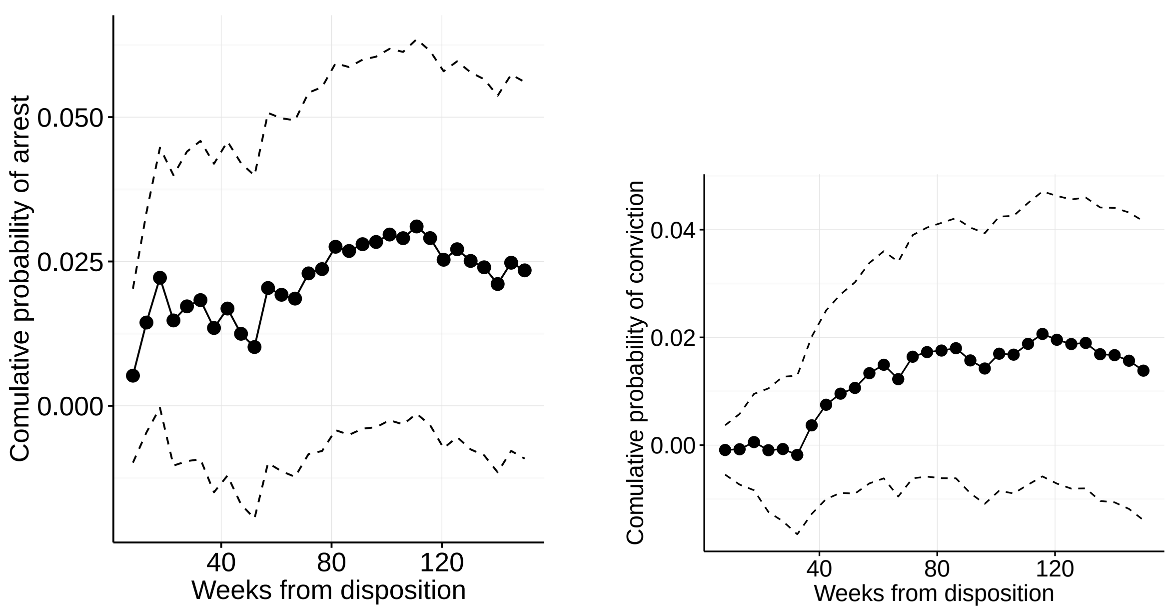
Notes: Each one of the plots above uses Monte-Carlo simulations to asses whether the observed F-statistic is likely under a mechanism which randomly assigned defendants across attorney types. Fischman (2011), Abrams et al. (2012) and Abrams and Fackler (2017) all used similar Monte-Carlo simulation procedures when assessing covariate balance and to correct finite sample coverage concerns with the asymptotic distribution of the conventional F-statistic. The red line shows the observed F-statistic and the histogram plots an approximation of the distribution of the F-statistic under a random assignment mechanism using 1,000 random re-labellings of defendants across attorney types. I randomly permuted/shuffled which defendants have been assigned to a PD relative to a CA, and then estimated the F-statistic for the null that all the coefficients are equal to zero. In the multiple defendant sample the permutations are done within a case. The number of re-labellings we use is 1,000 and it is similar to what is commonly used in the statistics literature. For example, Athey et al. (2018) and Anderson and Magruder (2017) use 1,000 draws/re-labellings; and Keele and Miratrix (2017) use 500 draws/re-labellings.

Table A.2: P-values of observed effects in Figure (1.5)

	asinh(Prison term)	Prison	Convicted
Female	0.468	0.575	0.690
Black	0.275	0.301	0.972
Hispanic	0.555	0.502	0.181
Prior incarceration	0.075	0.051	0.684
Prior conviction	0.032	0.005	0.250
Felony case	0.034	0.033	0.295

Notes: Each cell in the table reports the P-value of the observed effect (red triangular) in Table (1.5). The P-value is the number of times that a the estimated effect under a random permutation of treatment (black dots) was more extreme than the observed estimated effect.

Figure A.2: San Francisco: The relationship between initial assignment to a PD and involvement in the criminal justice system within a fixed period of time since disposition



A. Re-arrest with X weeks of disposition B. Re-conviction with X weeks of disposition

Notes: Each point in the figures is the estimated β coefficient from equation (1.2), where the outcome, Y_i, is a new arrest/conviction within a certain period of time from the date of disposition. The x-axis measures the time from disposition in weeks. The standard errors are cluster-robust at the case level. The recidivism measures are calculated only using new offenses/convictions in San Francisco.

Table A.3: San Francisco: The effect of having a PD vs. a CA on the case sentencing outcomes when controlling for attorney characteristics

	Initial PD effect			
	(1)	(2)	(3)	(4)
asinh(Prison term)	-0.118^{***}	-0.119^{***}	-0.101	-0.108
	(0.044)	(0.045)	(0.090)	(0.089)
Prison	-0.021^{***}	-0.021^{***}	-0.015	-0.016
	(0.008)	(0.008)	(0.016)	(0.016)
Convicted	-0.040^{***}	-0.046^{***}	-0.004	-0.012
	(0.014)	(0.014)	(0.025)	(0.025)
Case FE	Yes	Yes	Yes	Yes
Defendant controls	No	Yes	No	Yes
Attorney controls	No	No	Yes	Yes
Observations	6,703	6,703	6,703	6,703

Notes: Each cell in the table contains the coefficient on an indicator whether the defendant was initially assigned a PD or a CA. The standard errors are cluster-robust at the case level. Both incarceration and prison terms are measured in months. I approximate the Log(·) function using the asinh(·) function which is a common procedure when the outcome of interest is both skewed and has a mass at zero. The attorney characteristics include all the covariates in Table (1.4). The number of observations in this table is smaller than in Table (1.3), 6703 vs. 7164, since in some of the observations the attorney type was available but the attorney name was either not available or was partially listed.
*p<0.1; **p<0.05; ***p<0.01

Table A.4: San Francisco: Changes in attorney characteristics between first and terminating attorneys

	CA	PD	RE
Change attorney	0.207	0.521	0.161
Higher rank JD (US news)	0.010	0.080	0.006
Higher rank BA (US news)	0.079	0.232	0.060
Higher experience	0.163	0.354	0.142
Lower rank JD (US news)	0.011	0.089	0.014
Lower rank BA (US news)	0.091	0.214	0.070
Lower experience	0.100	0.167	0.080

Additional figures and tables from federal district courts

Figure A.3: Federal courts, the distribution of defendants across attorney types and over time, by filing year

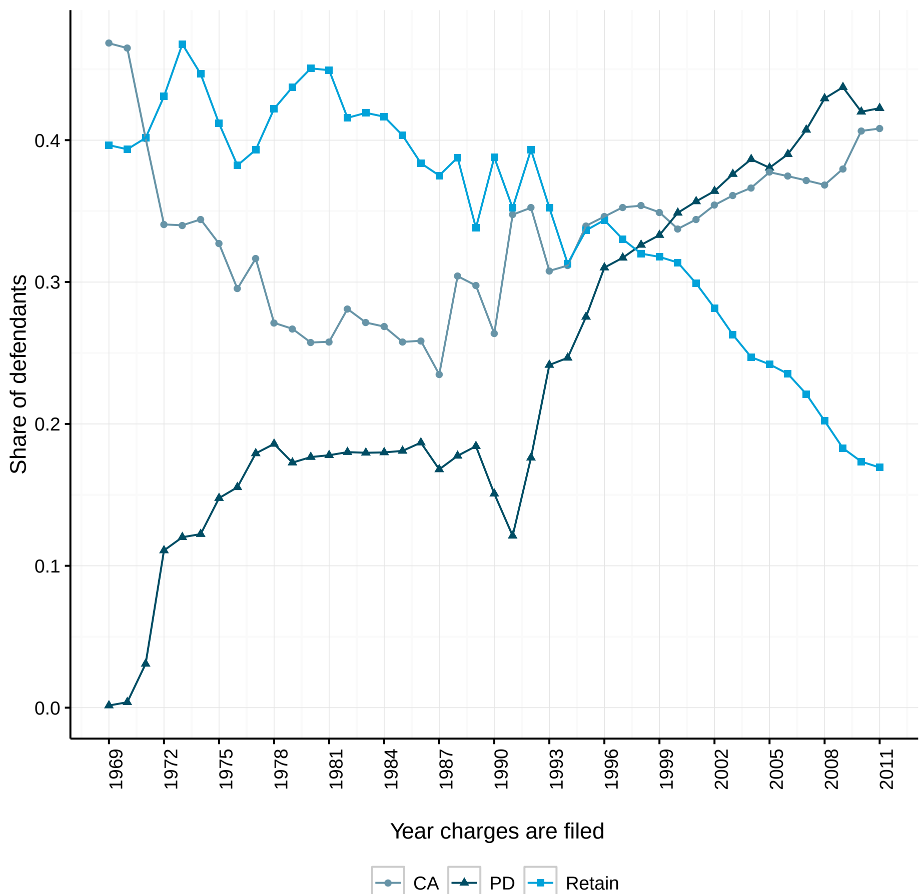
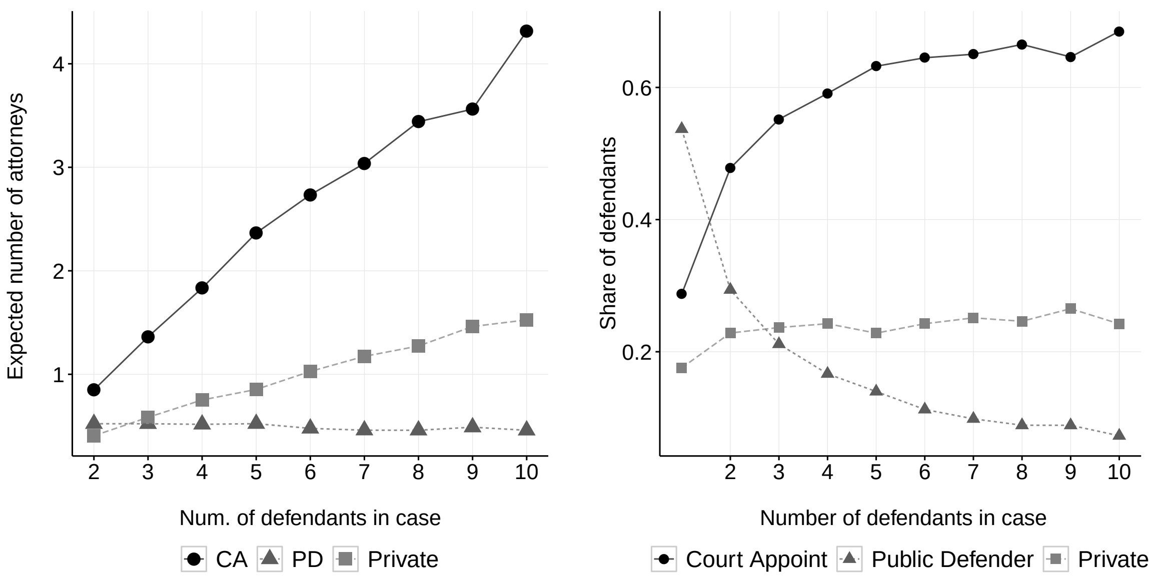


Figure A.4: Distribution of defendants across attorney types, by the num. of defendants in the case



A. Ave. num. of attorneys B. Distribution of def. across attorneys

Notes: The figure presents descriptive evidence on the distribution of defendants across attorney types in multiple defendant cases in federal district courts. See also the notes in Figure 1.2.

Table A.5: Descriptive statistics on indigent defendants in federal courts (1996 − 2014)

	All	Single def.	Multiple def.	Multiple def. (PD & CA)
	(1)	(2)	(3)	(4)
Prison term (months)	38.970	30.360	61.070	54.730
Prison	0.860	0.860	0.850	0.860
Conviction	0.950	0.960	0.940	0.940
Some plea	0.930	0.940	0.890	0.900
Trial	0.030	0.020	0.060	0.050
Acquittal	0	0	0.010	0.010
Predicted prison term	40.640	32.320	61.980	57.710
Predicted prison	0.850	0.850	0.860	0.870
Probation term	3.100	3.120	3.050	2.960
Predicted plea	0.920	0.930	0.900	0.900
Predicted trial	0.040	0.030	0.060	0.050
Predicted conviction	0.950	0.950	0.940	0.940
Predicted num. convictions	1.120	1.070	1.260	1.230
Predicted num. dismissed	0.710	0.430	1.420	1.360
Num. felony	1.680	1.310	2.610	2.520
Number of defendants	2.350	1	5.820	5.480
CA in case	0.510	0.350	0.930	1
PD in case	0.630	0.650	0.590	1
Observations	651,666	468,791	182,875	84,260

Notes: The table presents descriptive statistics for all criminal defendants in cases terminated in federal district courts from 1996 to 2014. The four columns refer to different sub-samples of the data. The column (1) makes no restrictions and includes all defendants. Column (2) restrict attention to individuals in single defendant cases, without any co-defendants. Column (3) Restrict the sample to individuals in cases that includes more than one defendant. Column (4) restrict the sample in column (3) to multiple defendant cases in which at least one defendant is represented by a PD and another by a CA. In each of these cases there are both defendants who are represented by a PD and a CA, which allows to conduct a within case comparison of attorney types. All the predicted variables are summary measures for the charges that have been filed against the defendant based on a Oaxaca decomposition. Appendix (A) describes how the predicted variables are constructed. Table (A.6) in Online Appendix (A) presents similar descriptive information for cases that have been terminated in federal district courts between 1970 to 1995.

Table A.6: Descriptive statistics on indigent defendants in federal courts ($1970 - 1995$)

	All	Single def.	Multiple def.	Multiple def. (PD & CA)
	(1)	(2)	(3)	(4)
Prison term (months)	27.600	21.800	37.260	35.340
Prison	0.520	0.500	0.570	0.590
Conviction	0.830	0.850	0.800	0.820
Plea	0.730	0.770	0.660	0.700
Trial	0.130	0.100	0.170	0.150
Acquittal	0.030	0.020	0.030	0.030
Predicted prison term	28.920	24.360	36.500	36.690
Predicted prison	0.530	0.500	0.580	0.600
Predicted probation term	67.650	71.240	61.670	50.380
Predicted plea	0.710	0.720	0.700	0.710
Predicted trial	0.140	0.130	0.160	0.160
Predicted conviction	0.820	0.820	0.830	0.830
Predicted dismissed	0.170	0.170	0.170	0.160
felony	0.470	0.420	0.570	0.690
Number of defendants	2.380	1	4.690	5.210
CA in case	0.720	0.600	0.900	1
PD in case	0.400	0.400	0.400	1
Observations	494,822	309,083	185,739	50,036

Notes: The table presents descriptive statistics for all criminal defendants in cases terminated in federal district courts from 1970 to 1995. The four columns refer to different sub-samples of the data. The column (1) makes no restrictions and includes all defendants. Column (2) restrict attention to individuals in single defendant cases, without any co-defendants. Column (3) Restrict the sample to individuals in cases that includes more than one defendant. Column (4) restrict the sample in column (3) to multiple defendant cases in which at least one defendant is represented by a PD and another by a CA. In each of these cases there are both defendants who are represented by a PD and a CA, which allows to conduct a within case comparison of attorney types. All the predicted variables are summary measures for the charges that have been filed against the defendant based on a Oaxaca decomposition. Appendix (A) describes how the predicted variables are constructed.

Table A.7: Federal courts, terminated cases 1996 − 2014: Difference in filed charges between defendants assigned to PD and CA across samples

	All indigent	All multiple	Multiple (PD & CA)	Multiple (PD & CA)
Num. felony	-0.450***	-0.121***	0.0688***	0.00470
	(0.00451)	(0.00801)	(0.00572)	(0.00648)
Num. misdemeanor	0.0142***	0.00378**	-0.00124	-0.000147
	(0.000981)	(0.00126)	(0.000737)	(0.000856)
Num. petty	0.0180***	0.00158**	-0.000487	0.0000695
	(0.000536)	(0.000541)	(0.000483)	(0.000524)
Predicted prison term	-9.086***	-4.809***	1.212***	0.0613
	(0.121)	(0.215)	(0.119)	(0.135)
Predicted prison	-0.0117***	-0.0114***	0.00192***	0.000173
	(0.000380)	(0.000566)	(0.000273)	(0.000320)
Predicted probation term	0.168***	0.341***	-0.0380***	-0.00321
	(0.0100)	(0.0174)	(0.00747)	(0.00887)
Predicted some plea	0.00697***	0.00195***	-0.000909***	-0.0000194
	(0.000130)	(0.000222)	(0.000133)	(0.000153)
Predicted trial	-0.00698***	-0.00294***	0.00163***	0.000114
	(0.000121)	(0.000230)	(0.000154)	(0.000174)
Predicted conviction	0.00105***	-0.000713***	0.000550***	0.0000968
	(0.0000811)	(0.000120)	(0.0000798)	(0.0000928)
Predicted mum. dismissed	-0.356***	-0.0941***	0.0478***	0.00356
	(0.00327)	(0.00581)	(0.00412)	(0.00468)
N	468,791	182,875	84,260	84,260
Def. Num. FE	No	No	No	Yes
Case FE	No	No	Yes	Yes
District FE	Yes	Yes	No	No
Year FE	Yes	Yes	No	No

Notes: Each cell in the table reports the coefficient of an indicator whether the defendant was initially assigned a PD or a CA. This is the β coefficient from estimating model (1.4). Each on of the columns reports estimates of β under a specification with different FEs. Standard errors in parenthesis are clustered at the case level.

*p<0.1; **p<0.05; ***p<0.01

Table A.8: The relationship between initial assignment to PD and being the defendant ranked with the highest predicted sentencing outcome (e.g., prison term, prison)

	All indigents	All multiple	Multiple defendants (PD & CA)		Co-defendants (PD & CA)
	(1)	(2)	(3)	(4)	(5)
Predicted prison term	0.125***	0.124***	0.0365***	0.00580	0.00330
	(0.00129)	(0.00270)	(0.00336)	(0.00381)	(0.00576)
Predicted prison	0.122***	0.117***	0.0338***	0.00588	0.00500
	(0.00124)	(0.00266)	(0.00336)	(0.00382)	(0.00578)
Predicted conviction	0.121***	0.112***	0.0266***	0.00598	0.00347
	(0.00126)	(0.00268)	(0.00336)	(0.00384)	(0.00579)
Predicted trial	0.126***	0.127***	0.0408***	0.0104**	0.00708
	(0.00129)	(0.00270)	(0.00336)	(0.00382)	(0.00577)
obs	651666	182875	84260	84260	35753
PositionFE	No	No	No	Yes	Yes
CaseFE	No	No	Yes	Yes	Yes
DistrictFE	Yes	Yes	No	No	No
YearFE	Yes	Yes	No	No	No

Notes: Each cell in the table reports the coefficient of an indicator whether the defendant was initially assigned a PD or a CA. This is the β coefficient from estimating model (1.4). Unlike Table (A.7), outcome represents an indicator for whether the defendant was ranked as facing the most severe charges according to a certain criterion, which varies by each one of the rows. For example, the cell in the first row and the second column, reports the different between defendants assigned a PD vs. a CA in the probability of being the defendant ranked with the longest expected prison term based on the severity of the charges. Standard errors in parenthesis are clustered at the case level.
*p<0.1; **p<0.05; ***p<0.01

Figure A.5: federal courts, validating the conflict of interest hypothesis $(1970 - 1995)$

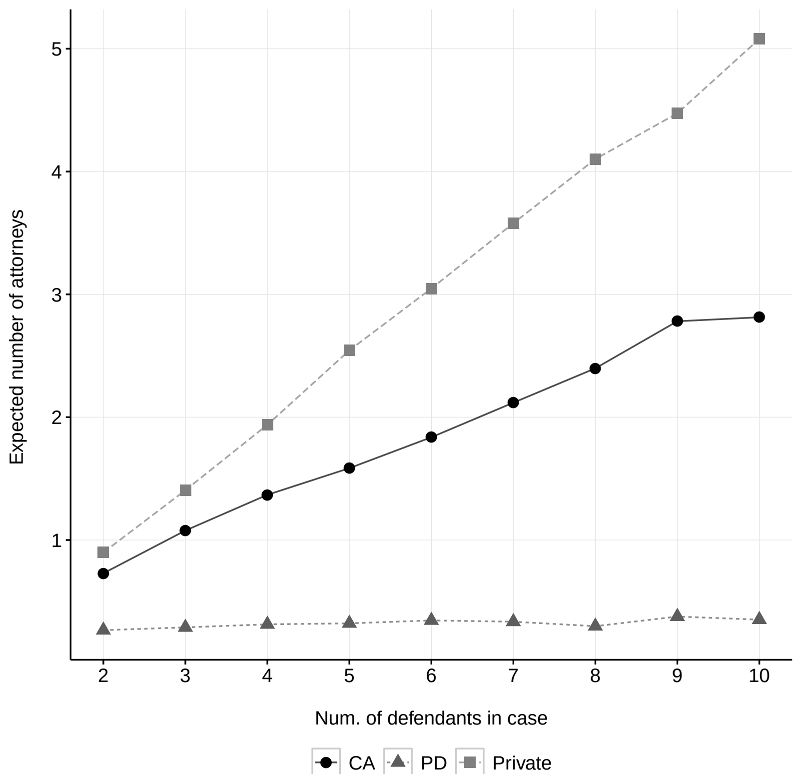
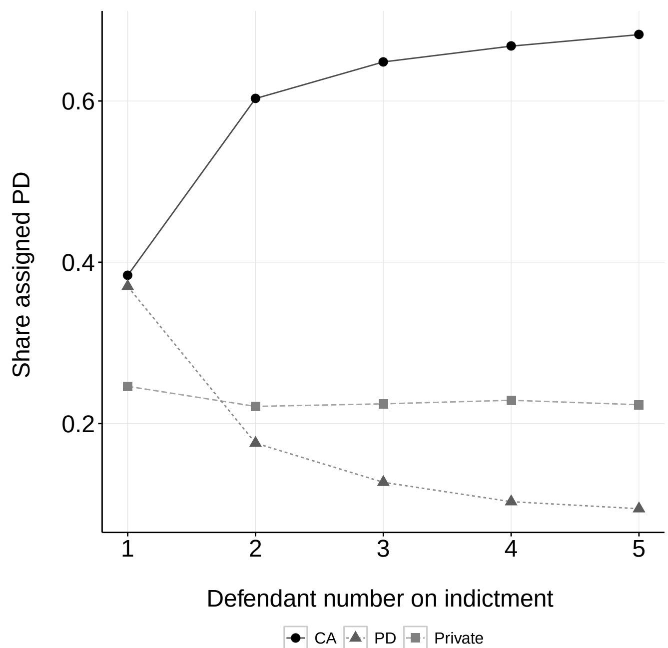
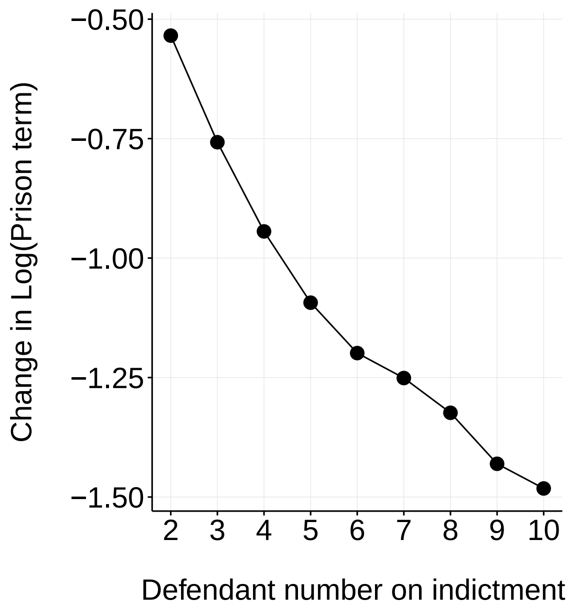


Figure A.6: Order on the indictment and the probability to be assigned PD among multiple defendant cases



Notes: The figure reports the distribution of defendants across attorney types by the number of the defendant on the indictment. For example, among defendants who are listed first on the indictment approximately 40% will be assigned to a PD, another 40% a CA, and the remaining 20% will be represented by a private attorney. The share of defendants who are represented by a PD is decreasing by the number of the defendant on the indictment.

Figure A.7: Order on the indictment and the length of incarceration



Notes: The figure reports the δ coefficients from equation (1.5). Each point is a fixed effect for a different position on the indictment, where the omitted category is the first defendant and all the coefficients report relative differences compare to the defendant that is listed first on the indictment.

Detecting random assignment of defendants across attorney types using a data driven procedure in federal district courts

Iyengar (2007) proposed a data-driven procedure to detect location-year pairs in which the assignment of defendants to attorney types—PD or CA—was done at random. A location-year pair is classified as using a random assignment mechanism if the covariates are not predictive of the treatment allocation. More specifically, consider the following model:

$$PD_i = X_i'\alpha + \gamma_{d(i)} + \eta_{t(i)} + \epsilon_i \tag{A.1}$$

and under random assignment of treatment the covariates should *not* be predictive of assignment to a PD:

$$H_0: \ \alpha = 0 \tag{A.2}$$

The procedure proposed by Iyengar (2007) is to conduct the hypothesis test in equation (A.2) in each district-year pair and if the F-statistic is below a certain threshold and is not statistically significant then to classify that district-year pair as using random assignment of defendants across attorney types. The above is a slight variation on the algorithm used by Iyengar (2007), since I am using ordinary-least-square instead of a Probit regression.

I find that no district passes a threshold of $F - stat < 0.5$ for two consecutive years; and 29 districts pass a threshold of $F - stat < 1$ for two consecutive years but only six pass this threshold for three consecutive years and no district passes the threshold for four consecutive years. It is not likely that districts frequently change the assignment procedure and therefore it is not clear whether the algorithm leads to false classification of districts as using random assignment or perhaps there is an over-rejection problem.

Next covariate balance is assessed in the sample of districts that passes the above procedure and are classified as using a random allocation procedure. The following econometric model is used to asses the covariate balance:

$$X_i = \beta \cdot PD_i + \gamma_{d(i)} + \eta_{t(i)} + e_i \tag{A.3}$$

where γ_d and η_t are district and filing year fixed effects. The β coefficient can be interpreted as the difference in means in characteristic X_i between defendants who have been assigned a PD relative to a CA within a given year and district.

Since we include district and filing year fixed effects the covariates should not be predictive of the attorney type—PD or CA—assignment. Table (A.9) reports the results for several different thresholds of the F-statistic. In column (1), all the balance tests look good; however, no district passes this threshold for two consecutive years. In column (2) there are significant imbalances and the differences increase as the $F - stat$ threshold is increased to 1.5 (column 3).

Table A.9: Tests for whether charge severity measures predict initial PD assignment among districts with imputed "random" allocation of PDs

	Initial assignment public defender		
	$F - stat < 0.5$	$F - stat < 1$	$F - stat < 1.5$
	(1)	(2)	(3)
Predicted prison term	−0.001	−0.0002	−0.0002*
	(0.0004)	(0.0001)	(0.0001)
Predicted prison	0.004	−0.190	−0.452***
	(0.281)	(0.121)	(0.082)
Predicted probation term	−0.002	−0.003	−0.009***
	(0.008)	(0.003)	(0.002)
Predicted plea	−0.046	1.252**	1.597***
	(1.043)	(0.599)	(0.433)
Predicted trial	0.306	1.518***	1.779***
	(0.814)	(0.486)	(0.362)
Predicted conviction	−0.116	−0.638	−0.464
	(1.047)	(0.590)	(0.419)
Predicted num. convictions	0.076	0.085***	0.080***
	(0.076)	(0.032)	(0.016)
Predicted num. dismissed	0.021	−0.048**	−0.103***
	(0.050)	(0.022)	(0.014)
Observations	2,197	21,500	56,721

Notes: The table includes also dummy variables for the number of felony charges the defendant is charged and have been removed from the table due to space limitations. Robust standard errors in parenthesis. The data includes only single defendant cases with at least one felony level charge.
*p<0.1; **p<0.05; ***p<0.01

The history of the right to appointed-counsel in the U.S.

The Sixth Amendment in the Bill of Rights states that "[i]n all criminal prosecutions, the accused shall enjoy the right... to have the Assistance of Counsel for his defense." However, the Constitution leaves open the question of what should happen when a defendant cannot afford to hire an attorney. In 1932, the Supreme Court ruled (*Powell v. Alabama*) that defendants charged with capital cases in state and federal courts who cannot afford an attorney have a constitutional right to have one appointed by the court. In 1938, the Supreme Court extended *Powell* and ruled that federal defendants in *all* felony criminal cases have a right to an appointed counsel (*Johnson v. Zerbst*). However in 1942, the Supreme Court decided in *Betts v. Brady* that *Johnson v. Zerbst* did not extend to defendants charged with non-capital cases at the state level. Although federal criminal defendants had the right to an appointed counsel since 1938, the question of who will compensate the appointed counsel remained open, and the provision of professional legal counsel to federal low-income defendants was limited (Judicial-Conference, 1952).

The Supreme Court established the right of indigent defendants to a court-appointed counsel in the 1960s. In 1963, the landmark ruling of the Supreme Court in *Gideon v. Wainwright* extended the limited scope of the *Powell* and *Johnson* decisions when it overturned *Betts v. Brady* by requiring states to provide a legal counsel to defendants facing any felony charges. In 1972, these rights were further extended to all criminal prosecutions that carry a sentence of imprisonment in *Argersinger v. Hamlin*. In *Scott v. Illinois*, the Supreme Court interpreted *Argersinger v. Hamlin* as referring only to a sentence of *actual* imprisonment. It determined that the criterion for whether a defendant is entitled to a court-appointed counsel is whether he was sentenced to an actual period of incarceration. In 2002, *Shelton vs. Alabama* extended the right to a court-appointed counsel also to defendants who are sentenced to a suspended sentence of incarceration (e.g., probation).[1]

In 1964, the Criminal Justice Act assured federal defendants professional legal counsel by establishing a federal indigent defense system financed by the court.[2] The CJA secured compensation for court-appointed attorneys and provided indigent defendants with funds for investigative and expert services to guarantee an adequate defense. In 1970, the law was amended to allow courts to establish federal public defender organizations (Prado, 1995; Haugh, 1966).[3]

[1]The Supreme Court mentioned two conditions under which a defendant facing a suspended sentence of imprisonment will not be eligible to a court-appointed counsel. The first is if the state offers an opportunity to re-litigate the guilt or innocent of the defendant in any future revocation proceedings. The second scenario is that the probation term cannot be revoked and replaced by actual imprisonment.

[2]In state courts, CAs are also referred to as conflict attorneys, assigned counsel, or panel attorneys. In federal courts, CA are commonly referred to as CJA attorneys in reference to the Criminal Justice Act of 1964, which established the federal indigent defense system. In this paper, I refer to court-appointed private attorneys as CA and use the initials CJA as a reference to the Criminal Justice Act of 1964.

[3]The Criminal Justice Act allows several districts to share the services of a single Public Defender Office, as long as their cumulative number of cases is at least 200.

Covariate indices for charge severity measures

To quantify the gaps in the severity of the filed charges between defendants that are assigned a PD and those assigned a CA, I consider a simple summary measure of the selection based on a Oaxaca decomposition. A trial outcome (e.g., incarceration length) Y_{ig} can be modelled by projecting it on a set of pre-trial charge characteristics:

$$Y_{ig} = X_{ig}\beta_g + \nu_g, \quad \text{where} \quad g = \text{PD, CA} \tag{A.4}$$

The coefficient vector β_g has a causal interpretation under certain conditions (Fortin, Lemieux, and Firpo, 2011), and the fitted values $X_g\hat{\beta}_g$ are independent of $\hat{\nu}_g$ by construction. The average difference in the trial outcome, $\bar{Y}_{\text{PD}}-\bar{Y}_{\text{CA}}$, between attorney types can be written as (Oaxaca, 1973),

$$\bar{Y}_{\text{PD}} - \bar{Y}_{\text{CA}} = \hat{\beta}_{\text{CA}}\left(\bar{X}_{\text{PD}} - \bar{X}_{\text{CA}}\right) + \left(\hat{\beta}_{\text{PD}} - \hat{\beta}_{\text{CA}}\right)\bar{X}_{\text{PD}} \tag{A.5}$$

The first element in (A.5), $\hat{\beta}_{\text{CA}}\left(\bar{X}_{\text{PD}} - \bar{X}_{\text{CA}}\right)$, is the average difference in charge characteristics re-weighted by the effect of each characteristic on the trial outcome among defendants who are represented by a CA. This term represents selection on observables and will be zero in a standard balance test when:

$$\bar{X}_{\text{PD}} = \bar{X}_{\text{CA}} \tag{A.6}$$

One can summarize the imbalance in initial charge characteristics by estimating the difference in covariate indices $X_i'\hat{\beta}_{\text{CA}}$ that reduces the dimension of the covariate vector X_i to a single dimensional index. The idea of summarizing imbalance by the covariates' relationship to the outcome surface has been proposed in the past by several different procedures (Bowers and Hansen, 2009; Card et al., 2015; Paetzold and Winner, 2016; Leacy and Stuart, 2014).

In San Francisco, I use the covariate index, $X_i'\hat{\beta}_{\text{PD}}$, which is based on estimating β using only defendants that have been assigned a PD. More specifically, I regress each case outcome on a vector of charge, case and defendant characteristics such as demographic characteristics, criminal history, charge severity (e.g., felony, misdemeanor). The main covariates are listed in Table (1.2) and Figure (1.3). In addition, I use SC and BCS codes which are 2-digit and 3-digit classifications of offenses to broader categories.[4]

In federal courts, I follow the procedure that as was described above and use the covariate index, $X_i'\hat{\beta}_{\text{CA}}$.[5] More specifically, I regress each case outcome on a vector of charges

[4]The classification is done by the California Department of Justice, `https://oag.ca.gov/law/code-tables`.

[5]In federal courts throughout the sample there exists a large fraction of defendants that have been assigned CAs; however, in San Francisco the fraction of the defendants that have been assigned CA is much smaller then those who have been assigned PD. In federal courts, in some districts there is no PD office for part of the period that is why in federal courts I use $\hat{\beta}_{\text{CA}}$ and in San Francisco $\hat{\beta}_{\text{PD}}$. If I use the same index in both the results are the same. This choice has no implication on the results reported in the paper.

characteristics such as indicators for the charges the defendants is facing based on a four-digit offense code of the Federal Administrative Office of the Courts.[6] The four-digit codes have many values and hence displaying balance tests for indicators of each of the offense codes is not feasible. This is one of the motivations to use the above dimension reduction. I also include indicators for the number of charges at each severity level (e.g., misdemeanor, felony). In federal district courts, I do not observed criminal history and demographic information about the defendant, unlike San Francisco in which this information is available.

[6] For more details on the four digit codes see https://www.fjc.gov/sites/default/files/idb/codebooks/Criminal%20Code%20Book%201996%20Forward_0.pdf

Monte Carlo simulations of exact inference vs. cluster-robust standard errors

Figure (1.4) suggest that conducting inference over the null that the attorney type has no effect using Fisherian inference, also known as exact/permutation inference, can have higher power to reject the null of no effect when it is false. To better understand the power of the two different methods of conducting inference, I conduct a simple Monte-Carlo simulation that examines the performance of the two procedures with respect to power in a data generating process that resembles to the observed data.

Consider the following data generating process of a constant treatment effect. I use the multiple defendants analysis sample and define the observed value of $asinh(\text{Prison term})_i$ as $Y_i(0)$. The potential outcome under the treatment regime is:

$$Y_i(1) = Y_i(0) + \tau \cdot \text{PD}_i \tag{A.7}$$

Next I randomly assign defendants to treatment regimes (PD_i) using randomization within a multiple defendant case. I use 16 different values of τ and for each one conduct 1,000 random assignments of defendants to treatment and for each such assignment calculate the observed value of Y_i:

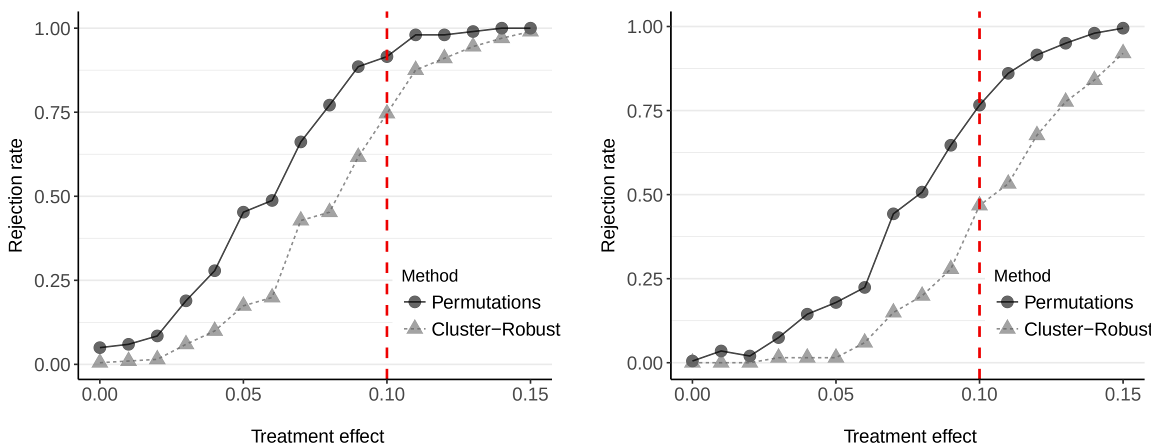
$$Y_i = Y_i(0) \cdot (1 - \text{PD}_i) + Y_i(1) \cdot \text{PD}_i \tag{A.8}$$

I then conduct inference over the null that τ is different than zero using both permutation inference and the cluster-robust standard errors. The estimator for τ is:

$$Y_i = \tau \text{PD}_i + \alpha_{j(i)} + e_i \tag{A.9}$$

and when using permutation inference the test statistic is the t-statistic of $\hat{\tau}$ divided by the cluster-robust standard error both are calculated in each random permutation of treatment assignment. Figure (A.8) reports the simulation results in terms of the rejection rate of the two procedures for each one of the values of τ. For higher values of τ both procedures reject the null in a higher rate, however, the permutation inference procedure seems to have a higher rejection rate for every τ and stochastically dominates the cluster-robust inference in terms of power. This Monte-Carlo simulation is specific to this data application and should not be used to make general claims on the efficiency of permutation inference relative to regular inference based on cluster-robust standard errors.

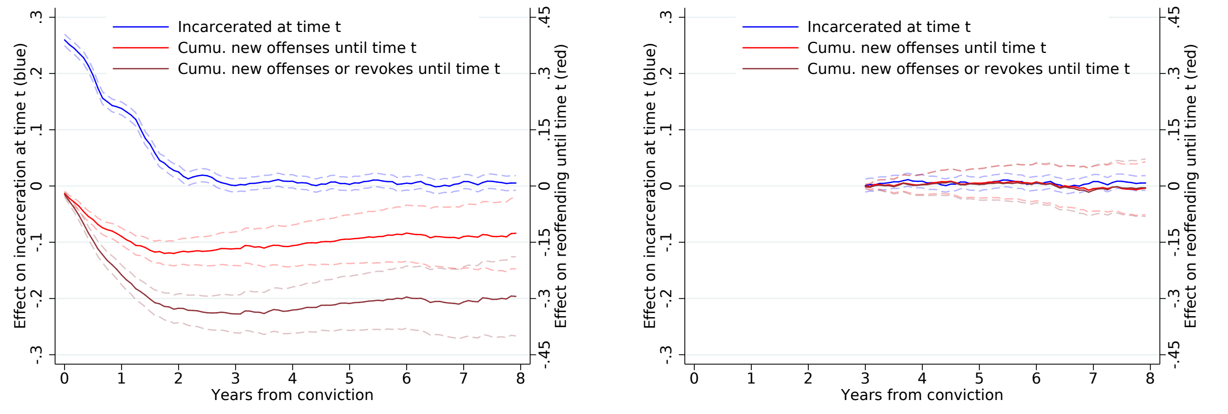
Figure A.8: **The rejection rate using permutation inference compare to the standard cluster-robust standard errors**



A. Rejection rate at 5% confidence level   B. Rejection rate at 1% confidence level

Notes: The clustering, in the cluster-robust standard errors, is performed at the case level. This is the same level in which the randomization is conducted in the permutation inference calculations.

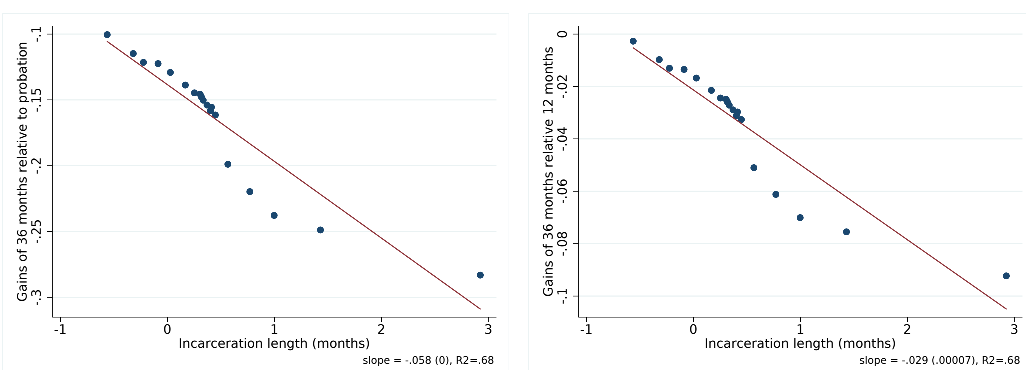
Figure A.9: Reduced form estimates of *cumulative* reoffending up to period t from conviction



(a) Cumulative reoffending until period t (b) Cumulative reoffending from 3 years until period t

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) 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 the cumulative number of new offenses committed until 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 same estimated effect only now the cumulative measures start only after three years from conviction, when the instruments no longer predict incapacitation status, and the maroon color line (right y-axis) the estimates when also including probation revocations as offending. Standard errors are clustered by individual.

Figure A.10: Selection into incarceration duration based on gains (reduction in reoffending due to incarceration exposure)



Gains: $D = 36$ relative to $D = 0$ Gains: $D = 36$ relative to $D = 12$

Notes: This figure shows binscatters of the correlation between assignment to longer spells of incarceration and two different types of measures. First, levels. Second, gains.

Table A.10: 2SLS estimates of effect of months of incarceration on various reoffending outcomes within one year of sentencing

	Measure of crime					
	(1) Re-incar	(2) Any new offense	(3) Felony	(4) Assault	(5) Property	(6) Drug
Months incap	-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
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: Dependent variable is an indicator for any charges (or conviction) recorded in the AOC (or DPS) data within one 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).

I Failure Functions and IV Estimators

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

$$\begin{aligned} Y_{i,t} &= Y_{i,t} + y_{i,t} \cdot (1 - Y_{i,t}) \\ Y_{i,t} &= y_{i,t} \end{aligned} \tag{A.10}$$

The recursive formulation implies that analogous Wald estimates of the failure function for t periods from conviction result in:

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

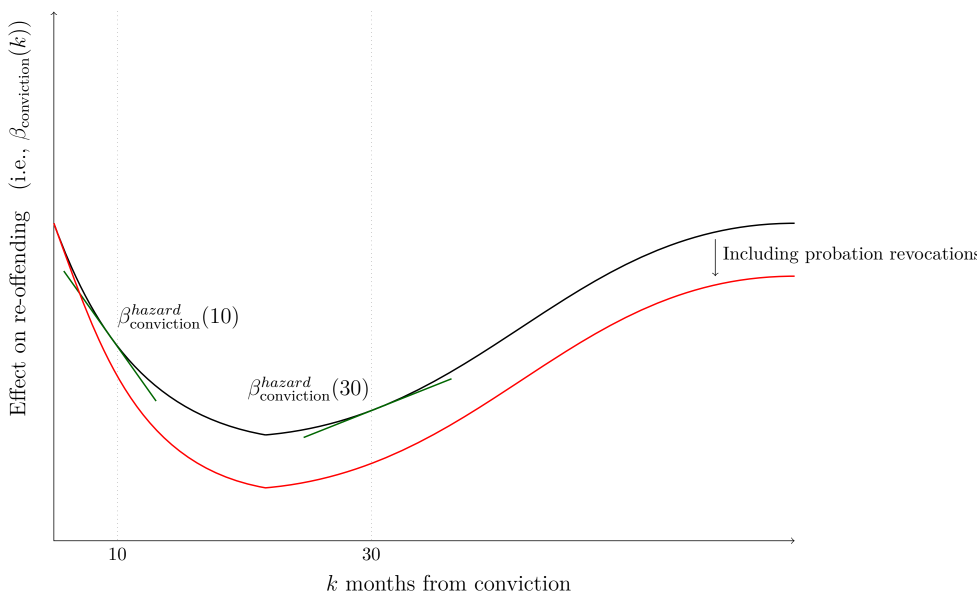
where the above derivations also hold for Wald estimates of the effects of incarceration on re-offending from the date of release.

The second term in equation (A.11) 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 $\beta^{\text{hazard}}_{\text{conviction}}(t)$:

$$\beta_{\text{conviction}}(t) = \sum_{l=1}^{t} \beta^{\text{hazard}}_{\text{conviction}}(l) \tag{A.11}$$

In what follows, 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 $\beta^{\text{hazard}}_{\text{conviction}}(t)$. Figure A.11 illustrate how estimates of $\beta_{\text{conviction}}(t)$ will look like. In the initial periods from conviction the incapacitation effect will dominate and the difference in re-offending rates between the treated and control units will be increasing. Once individuals who have been initially incarcerated are released the difference in offending rates either stabilizes or starts to diminish until it reaches a stable value. The tangent green lines indicate the slope of the $\beta_{\text{conviction}}(t)$ function which is equal to $\beta^{\text{hazard}}_{\text{conviction}}(t)$. The red line indicates the impact of counting probation revocations as re-offending and is discussed further in section II.

Figure A.11: Illustration of IV estimates of incarceration on re-offending within t months from conviction



II 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 re-offending 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 re-offending from the date an individuals is back in the community and is at-risk to re-offend. The estimand that is now recovered by 2SLS, without adjusting for any time-varying controls, is

$$\beta_{\text{at-risk}}(t) \equiv \frac{\mathbb{E}\left[Y_{i,t+D_i}|Z_i=1\right]-\mathbb{E}\left[Y_{i,t+D_i}|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]} \tag{A.12}$$
$$=\sum_{j=1}^{\bar{D}}\omega_j\mathbb{E}\left[Y_{i,t+d}(d)-Y_{i,t+d-1}(d-1)|D_i(1)\geq d>D_i(0)\right]$$

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

$$Y_{i,t+D_i}=Z_i\left[\sum_{d=1}^{\bar{D}}Y_{i,t+d}(d)\left(\lambda_{i,d}(1)-\lambda_{i,d+1}(1)\right)\right]+(1-Z_i)\left[\sum_{d=1}^{\bar{D}}Y_{i,t+d}(d)\left(\lambda_{i,d}(0)-\lambda_{i,d+1}(0)\right)\right] \tag{A.13}$$

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 re-offending is measured based on the endogenous treatment of interest D_i. Another concern is that the incarcerated and on-incarcerated offenders will now vary in observable and unobservable time-varying factors. For example, the age of the offender at time zero, i.e., the point in time that we start to measure re-offending.

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

$$\mathbb{E}\left[Y_{i,t+D_i}\right]=X_i'\alpha+W_{i,D_i}'\gamma+\sum_{i=d}^{\bar{D}}\gamma_d\mathbb{I}\left\{D_i\geq d\right\} \tag{A.14}$$

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

$$\sum_{j=1}^{\bar{D}}\omega_j\gamma_d+\sum_{j=1}^{\bar{D}}\omega_j\left(W_{i,d}-W_{i,d-1}\right)'\gamma \tag{A.15}$$

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

$$\sum_{j=1}^{\bar{D}}\tilde{\omega}_j\gamma_d,\quad \tilde{\omega}_j=\frac{\Pr(D_i\geq j)\mathbb{E}\left[\xi_i|D_i\geq j\right]}{\sum_{l=1}^{\bar{D}}Pr(D_i\geq l)\mathbb{E}\left[\xi_i|D_i\geq l\right]} \tag{A.16}$$

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 (A.16) is a linear combination of causal effects; however, the $\tilde{\omega}_j$ weights can have negative values—rolling 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:

$$\sum_{j=1}^{\bar{D}} \frac{\Pr(D_i \geq j)\mathbb{E}\left[\xi_i \gamma_{i,d} | D_i \geq j\right]}{\sum_{l=1}^{\bar{D}} Pr(D_i \geq l)\mathbb{E}\left[\xi_i \gamma_{i,d} | D_i \geq l\right]} \tag{A.17}$$

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 3) that makes additional parametric restriction but provides a semi-structural framework to identify the behavioral effects of incarceration.

III Independent competing risks

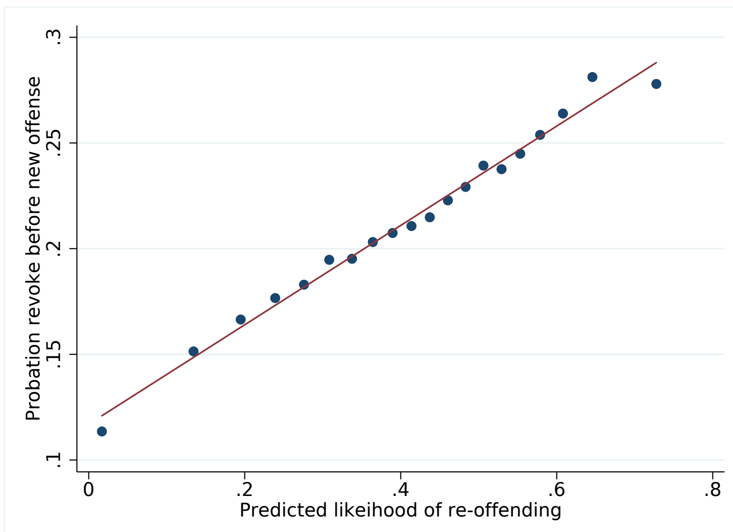
Reincarceration due to technical probation revocations can bias incarceration effects estimates in two ways. First, if revocations mask genuine criminal activity, not counting them as reoffending may artificially deflate reoffending rates in the probation (and thus control) population. Second, even if 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 A.12 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.[7]

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 A.13, 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 for a variety of types of new offenses (Appendix Table A.11). For example, the effect of a year of incarceration increases from 2.59 p.p. (↓37.6%) to -3.42 p.p. (↓47.1%). YTS: I'm not sure whether to argue

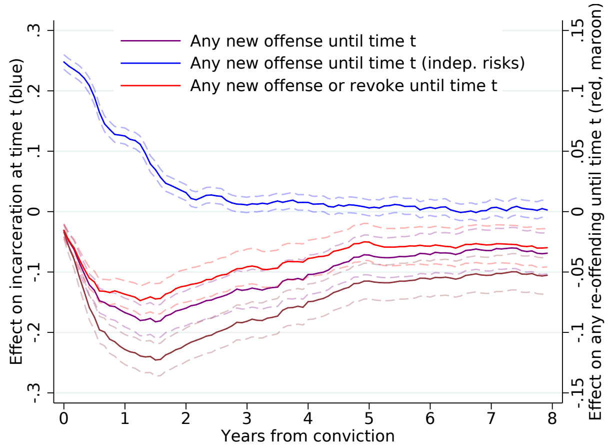
[7] It is important to note that this is a descriptive correlation only. We estimated an OLS model of committing a new offense within three years from at-risk on control variables that include criminal history and demographic information (e.g., age) using only individuals who have been incarcerated and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

Figure A.12: 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 (2.7), i.e., trends in prior record points within a prior record level (columns) and an offense felony class (rows).

Figure A.13: 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. Standard errors are clustered by individual. See also the notes in Figure 2.6 for further details on the estimation.

Table A.11: 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)	(2)	(3)	(4)	(5)	(6)
	Re-incar	Any new offense	Felony	Assault	Property	Drug
Months incap	-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
One year effect in %	-0.381	-0.279	-0.285	-0.471	-0.281	-0.242
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F (excluded-inst)	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.

Table A.12: Estimates of the effect of incarceration on reoffending from sentencing using charged vs. convicted offense class

	New offense			New offense of revoke		
	(1)	(2)	(3)	(4)	(5)	
	Arraigned	Charged	Convicted	Arraigned		
Months incarcerated	-0.00959***	-0.00960***	-0.00923***	-0.0146***	-0.0146***	-0.0144***
	(0.00165)	(0.00164)	(0.00102)	(0.00166)	(0.00166)	(0.00104)
N	363360	363360	363360	363360	363360	363360
Dep. var. mean	0.428	0.428	0.428	0.544	0.544	0.544

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.

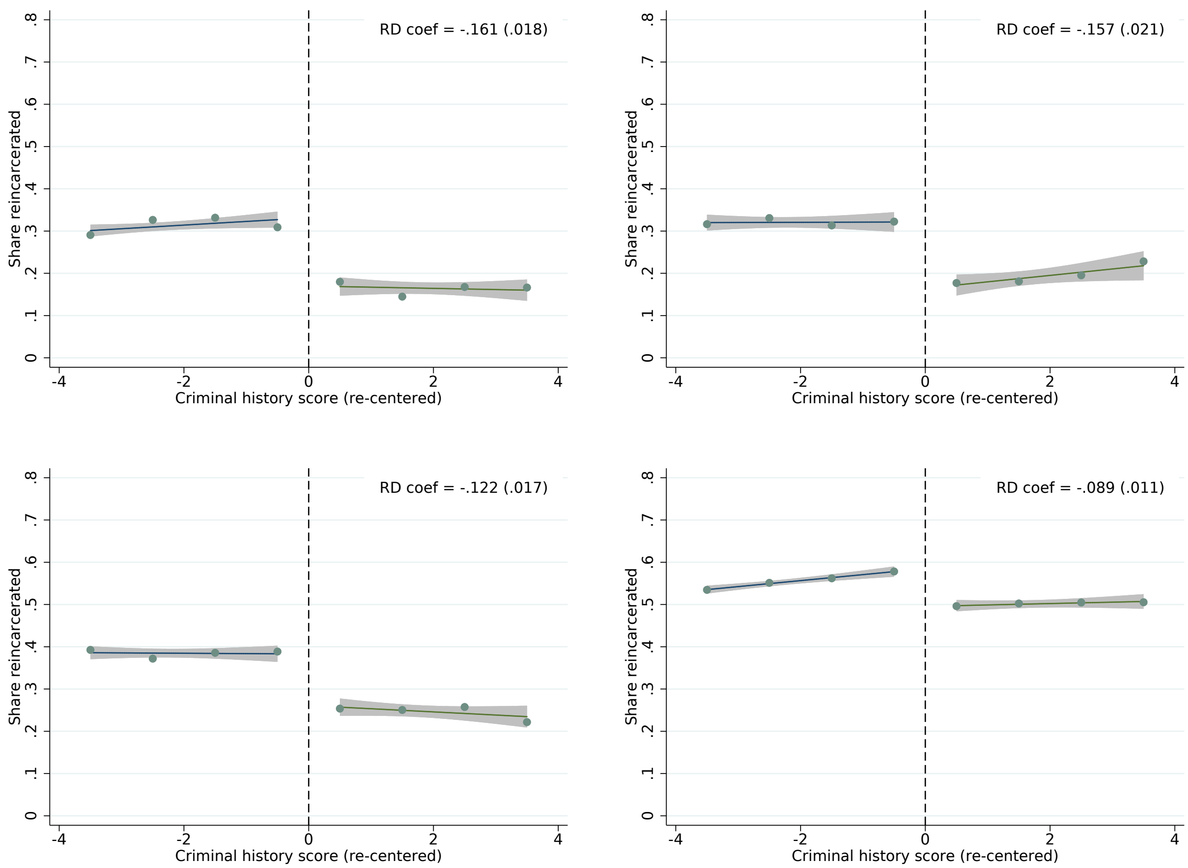
IV Heterogeneity by discontinuity

In this appendix, we explore heterogeneity by felony class and report estimates of reduced form figures that are analogous to Figure 2.6 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 (2.1)).

Appendix Figures A.19 and A.20 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.

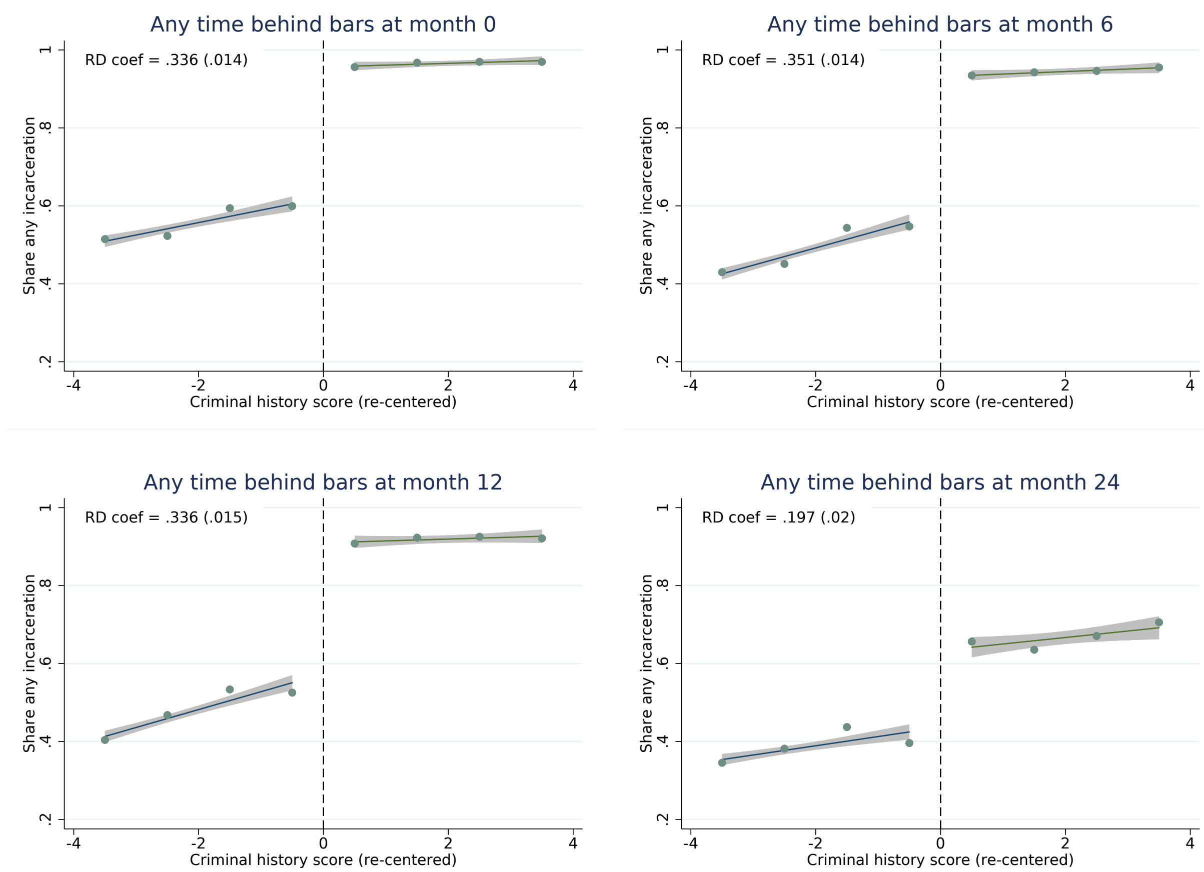
Panel (b) plots shows that although there is substantial heterogeneity in the magnitude of the incapacitation effects, the impacts on any reoffending in the long term show either a zero effect (e.g., class I) or permanent reduction in some classes (e.g., E or F). It is interesting to note that the reduced 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 A.14: Share reincarcerated within three years of conviction (offenses from felony classes E,G,H, and I)



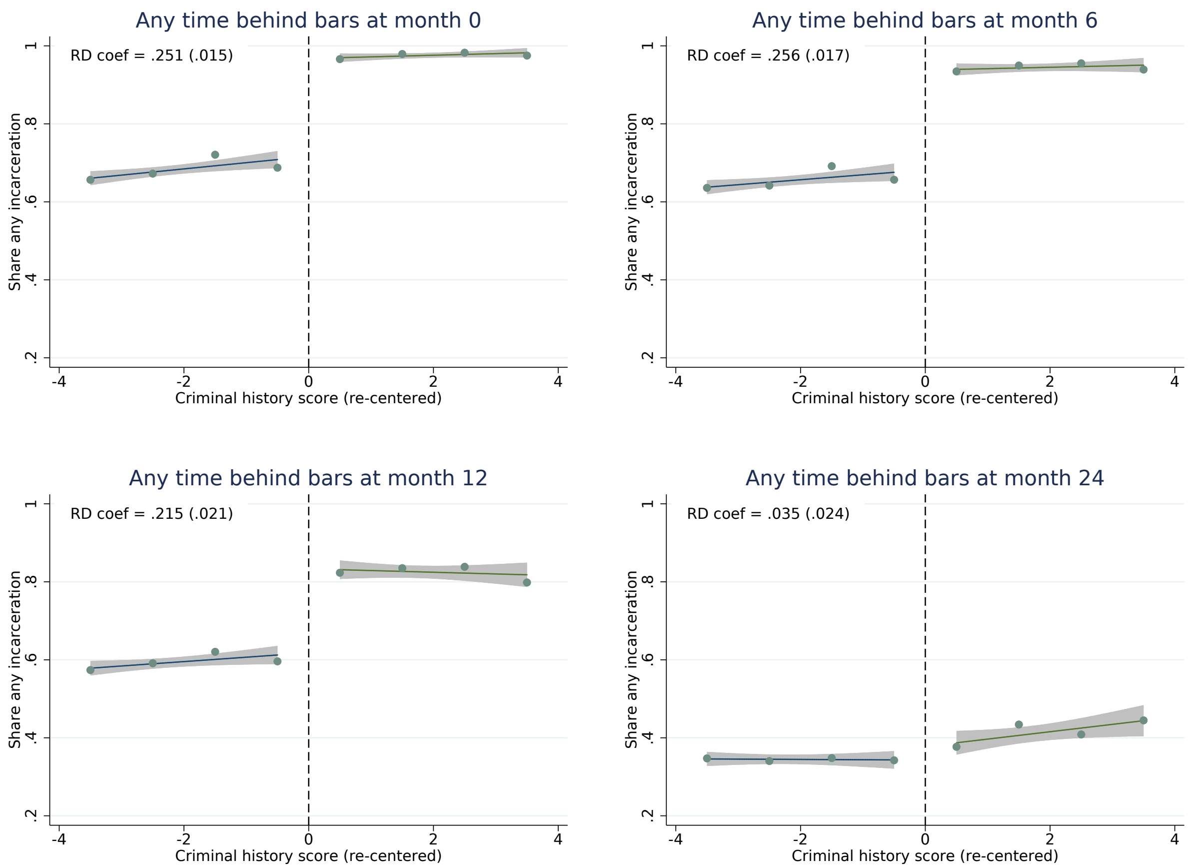
Notes: See notes to Figure 2.4.

Figure A.15: Dynamic differences in incarceration status *at* a given month after conviction (class E felony offenses)



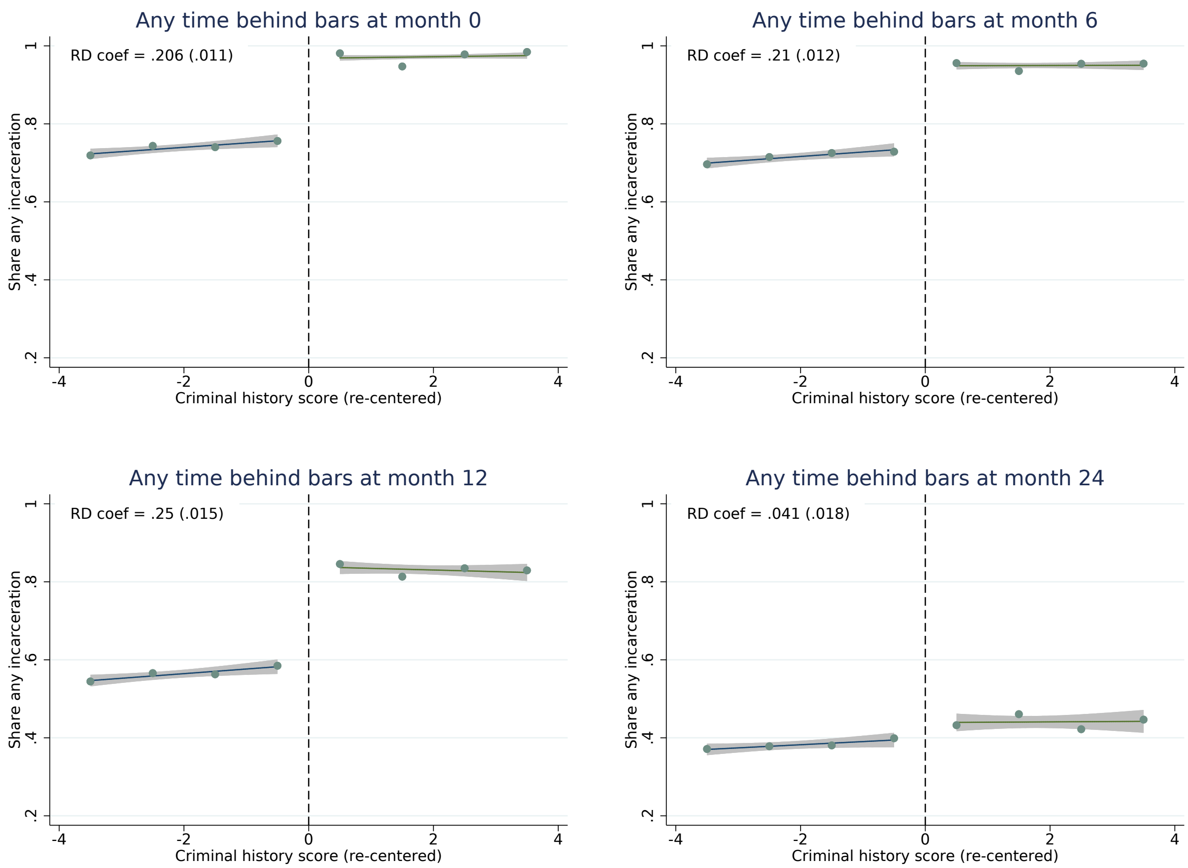
Notes: This figure shows reduced form RD estimates of being to the right of the punishment type discontinuity in class E on the likelihood of being incapacitated behind bars at month t after the date of conviction. The x-axis shows the recentered value of prior record points. The y-axis reports the share 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.

Figure A.16: Dynamic differences in incarceration status *at* a given month after conviction (class G felony offenses)



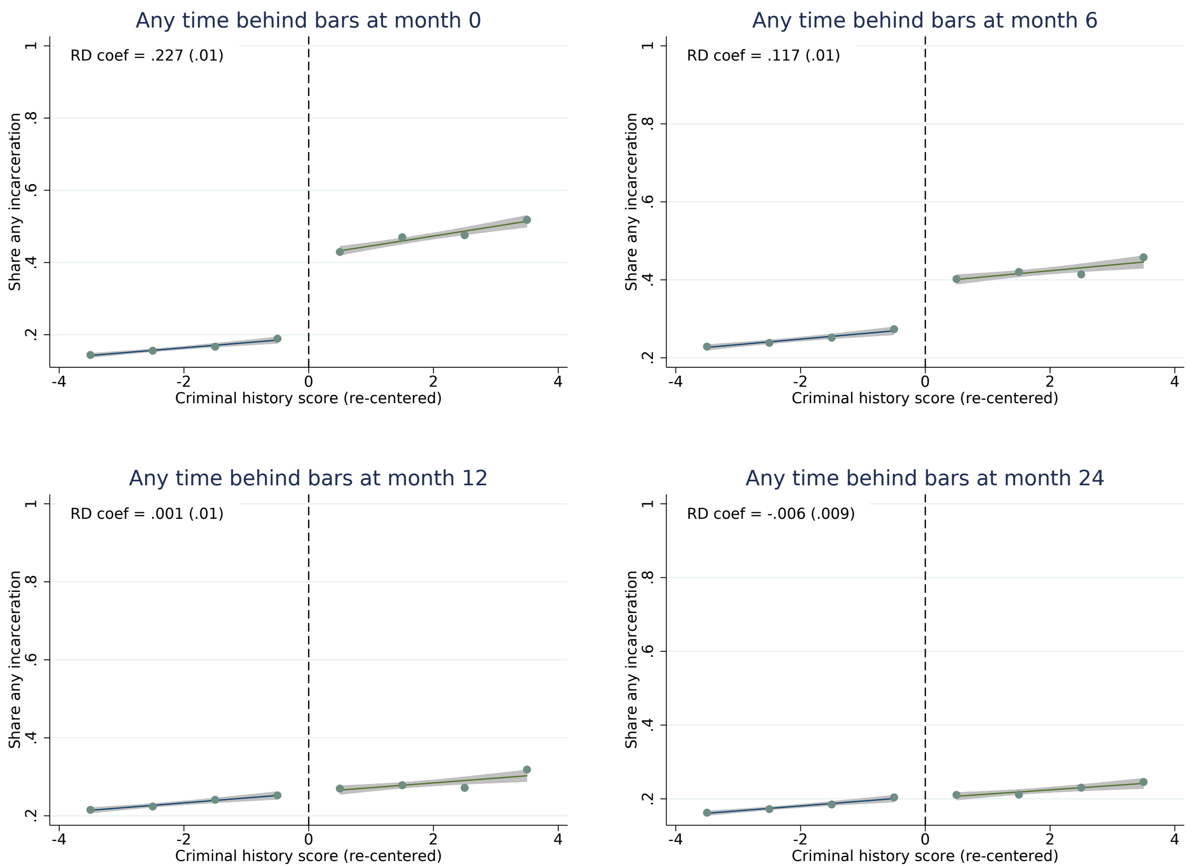
Notes: See notes to Figure A.16

Figure A.17: Dynamic differences in incarceration status *at* a given month after conviction (class H felony offenses)



Notes: See notes to Figure A.17

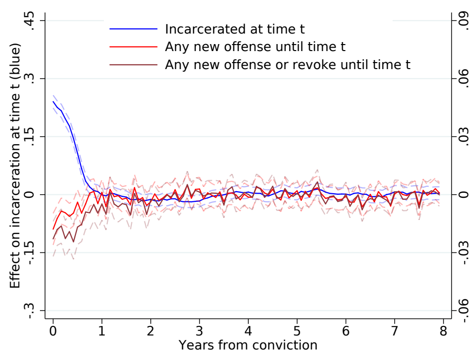
Figure A.18: Dynamic differences in incarceration status *at* a given month after conviction (class I felony offenses)



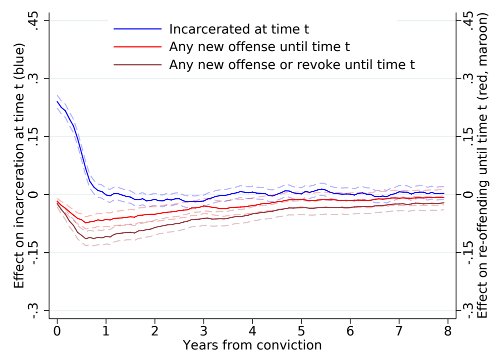
Notes: See notes to Figure A.18

Figure A.19: Reduced form estimates of reoffending *at* period t from conviction and also estimates of *any* reoffending up to period t from conviction

Class I

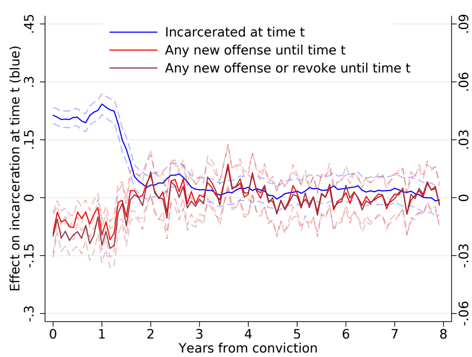


(a) reoffending at period t from conviction

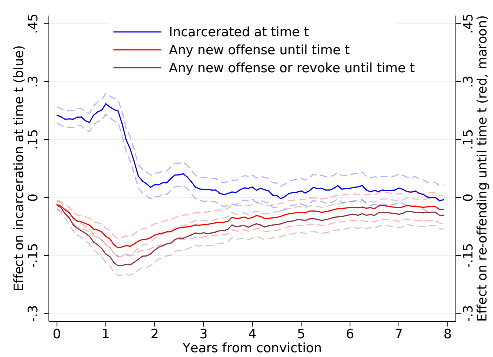


(b) Any reoffending until period t from conviction

Class H



(a) reoffending at period t from conviction

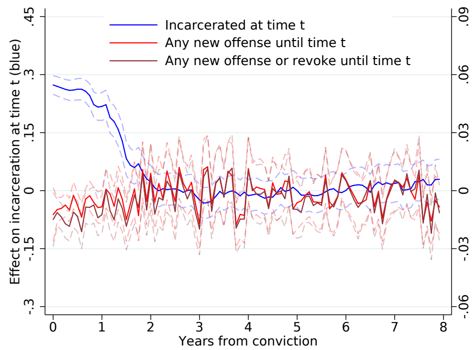


(b) Any reoffending until period t from conviction

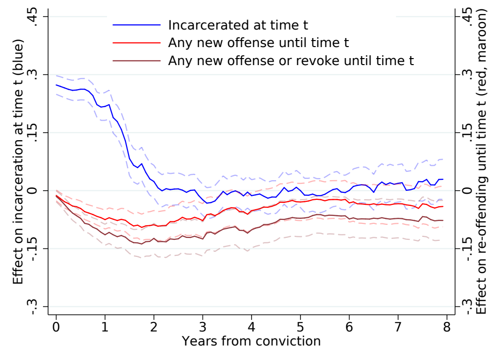
Notes: This figures shows reduced form estimates of being to the right of a punishment type discontinuity on several different outcomes of interest. All outcomes/measures are with respect to the conviction date.

Figure A.20: Reduced form estimates of reoffending *at* period t from conviction and also estimates of *any* reoffending up to period t from conviction

Class G

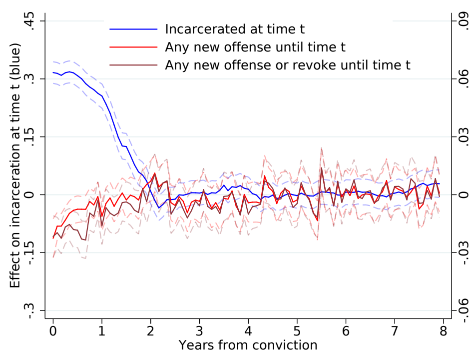


(a) reoffending at period t from conviction

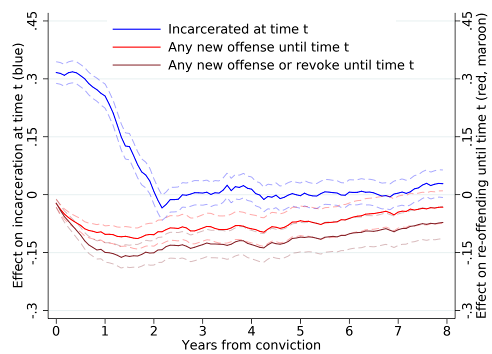


(b) Any reoffending until period t from conviction

Class F

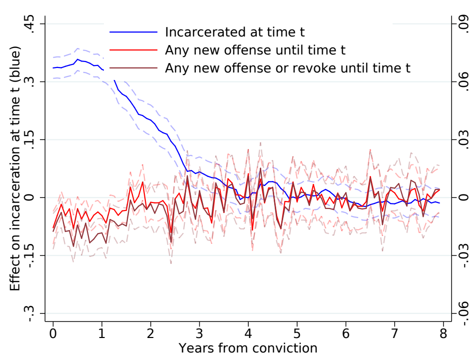


(a) reoffending at period t from conviction

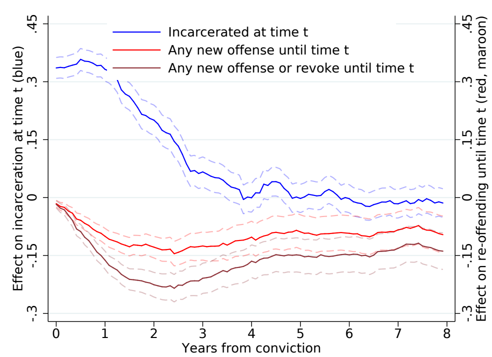


(b) Any reoffending until period t from conviction

Class E



(a) reoffending at period t from conviction



(b) Any reoffending until period t from conviction

Notes: See notes of above Figure A.19.