Conviction, Incarceration, and Policy Effects in the Criminal Justice System[*]

Vishal Kamat,[†] Samuel Norris,[‡] Matthew Pecenco[§]

March 27, 2024

Abstract

The criminal justice system affects millions of Americans through criminal convictions and incarceration. In this paper, we introduce a new method for credibly estimating the effects of both conviction and incarceration using randomly assigned judges as instruments for treatment. Misdemeanor convictions, especially for defendants with a shorter criminal record, cause an increase in the number of new offenses committed over the following five years. Incarceration on more serious felony charges, in contrast, reduces recidivism during the period of incapacitation, but has no effect after release. Our method allows the researcher to isolate specific treatment effects of interest as well as estimate the effect of broader policies; we find that courts could reduce crime by dismissing marginal charges against defendants accused of misdemeanors, with larger reductions among first-time defendants and those facing more serious charges.

[*] We thank Amanda Agan, Bocar Ba, Jason Baron, Peter Hull, Yupei Ma, Steve Mello, Paul Goldsmith-Pinkham, Vítor Possebom, Marit Rehavi, Ignacio Rodriguez Hurtado, Alexander Torgovitsky, and seminar audiences at Boston University, Brown, Duke, Interactions, Oregon, UC Santa Barbara, UC Irvine, UC Davis, Stanford, SOLE, SEA, and TxECW for their feedback. Diti Jain and Alison Lodermeier provided outstanding research assistance.

[†] Toulouse School of Economics, vishal.kamat@tse-fr.eu

[‡] University of British Columbia, sam.norris@ubc.ca

[§] Brown University. matthew_pecenco@brown.edu

Decision-makers, such as judges and doctors, can profoundly affect individual outcomes through the choices they make. As a result, quasi-randomly assigned decision-makers are widely used as instruments to estimate the causal effect of important policies and treatments as varied as incarceration, disability insurance, and hospital quality (Kling, 2006; Maestas, Mullen and Strand, 2013; Doyle et al., 2015). The majority of prior work has used these *examiner designs* to estimate the effect of a single treatment, such as incarcerating a defendant, and interpreted their analysis as identifying the heterogeneous effects of individuals who are marginal to a policy change (e.g., Aizer and Doyle Jr, 2015). However, examiners often make choices over more than two options—for example, whether to acquit, convict but not incarcerate, or convict and incarcerate—which threatens the validity and interpretability of this design (Heckman and Urzua, 2010).

In this paper, we develop a new framework to analyze settings with multiple observed treatments and randomly assigned examiners. We first study the common approach of instrumenting only for the treatment of interest and show that the resulting 2SLS estimand is interpretable only under restrictive assumptions on treatment effect heterogeneity or on how examiners' decisions might differ. Unlike other institutional features required in this design, such as random assignment, these restrictions are difficult to verify and researchers may not be confident that they hold exactly. Incorrect assumptions can lead to inconsistent or even wrong-signed estimates. To credibly account for these issues, we study a model of multiple treatment assignment that imposes only weak assumptions on decision patterns across examiners, and which can be translated into a tractable structural form for estimation. We show how our approach allows the researcher to consistently estimate traditional IV target parameters, decompose them into margin-specific subcomponents, and investigate a wide variety of policy-relevant treatment effects (Heckman and Vytlacil, 2005). The cost of our weaker assumptions on examiner behavior is that the target parameters are only partially identified, although we find that the bounds are highly informative.

We use this new framework to analyze the effects of conviction and incarceration in the US criminal justice system. Both of these criminal sanctions are common—in each year, approximately 9 million people are convicted and 1.7 million incarcerated—but they are typically studied separately and in different settings to avoid the conceptual issues arising from multiple treatments.[1] Convictions are usually studied in the context of minor charges where incarceration almost never occurs (Agan, Doleac and Harvey, 2022), while research on incarceration focuses on more serious offenses where the comparison group is almost always convicted (Garin et al., 2023). Consequently, little is known about the impacts of these sanctions for the millions of defendants outside of these relatively small groups or whether results from disparate environments can be replicated in a common setting. It also complicates efforts to understand the effects of potential policy changes, since most reforms would affect multiple margins; for example, raising the standards for a conviction would also reduce the incarceration rate.

[1] There are 13 million misdemeanor and 5 million felony cases filed each year in state courts (Ostrom et al., 2020). We use conviction rates from Chien (2020) to impute conviction numbers. BJS (2020) reports 530,000 felony incarcerations in 2020; we impute misdemeanor incarcerations using the incarceration rate in our sample (10%) and the Ostrom et al. (2020) case numbers.

We have three main findings. First, we find very different effects of conviction and incarceration. Incarceration reduces future crime through incapacitation, with felony incarceration reducing recidivism during the period the defendant is still in prison and misdemeanor incarceration (which on average results in only 30 days in jail) having a precisely estimated null effect. However, we find that misdemeanor convictions result in *higher* levels of future crime and—consistent with a key role for criminal records—these effects are larger for defendants without serious prior convictions.

Second, we investigate a series of possible criminal justice reforms that incorporate changes along both conviction and incarceration margins and show there is scope for policies to reduce punitiveness while improving public safety. Increasing judges' leniency in determining guilt for misdemeanor defendants would reduce future crime and involvement with the criminal justice system. These gains disproportionately accrue to defendants accused of more serious misdemeanor offenses, so policies focusing on this group would be even more beneficial.

Finally, in this setting 2SLS does not always reliably estimate the treatment effects of interest: a researcher instrumenting for conviction with judge assignment would dramatically overstate the increases in crime that result from conviction. We do, however, find that the exclusion and monotonicity violations are small when instrumenting for incarceration.

We begin by examining what 2SLS allows us to learn about defendant outcomes under conviction and incarceration, relative to dismissal. The standard approach in the literature is to instrument for a single treatment of interest with judge assignment. We decompose the 2SLS estimand into its constituent effects and show that it will typically not be interpretable as a treatment effect because of the presence of exclusion violations: when judges vary in multiple margins of treatment, it is not typically possible to attribute the variation in outcomes across judges to a single margin. The most common solution is to instrument for both treatments simultaneously using 2SLS; however, as is well-known from prior work, the resulting estimand will typically be interpretable only if the effect of each treatment is the same for all complier groups.

These findings motivate us to develop a model of examiner decisions over more than two potential choices that we can use to credibly recover treatment effects. The challenge is that our institutional setting requires a high degree of flexibility in response patterns. In particular, the law requires that judges consider different, unrelated factors when deciding whether to convict a defendant at trial and whether to incarcerate a guilty defendant during the sentencing process. It is likely that judges evaluate each of these different dimensions of choice in a heterogeneous way, which leads to specific response patterns such as pairs of judges with one set of compliers that move from conviction to incarceration, and another set that moves from incarceration to dismissal.

To account for this needed flexibility, we use a *latent monotonicity* assumption that is motivated by the institutional structure and accommodates the two-way flows of compliers prevalent in our setting. We show that latent monotonicity nests the compliance groups available in ordered monotonicity (Angrist and Imbens, 1995), unordered monotonicity (Heckman and Pinto, 2018), and single-index choice (Heckman, Urzua and Vytlacil, 2006), three models that are commonly used in settings with discrete instruments. The cost of this flexibility is that we are required to develop a novel identification approach. We translate our assumptions on potential outcomes into an equivalent threshold model of judge decisions and demonstrate that the parameters of this selection model—and thus, the size and even existence of each compliance group—are not identified from the data, making it unclear how one could estimate treatment effects.

One approach under latent monotonicity has been to condition on judge conviction propensities and estimate only the effect of incarceration relative to conviction (Arteaga, 2021).[2] We show how to go beyond this and estimate the relative effects of all three treatments. Our insight is that for each admissible selection model that is consistent with the observable judge treatment propensities, one can estimate marginal treatment response functions via simple linear regression and aggregate them up to recover a wide variety of 2SLS-weighted and policy-relevant treatment effects (Brinch, Mogstad and Wiswall, 2017; Heckman and Vytlacil, 1999, 2005). Under a semiparametric assumption on the primitives that drive selection, we can tractably search across all of these admissible selection patterns. Since the true underlying selection pattern is not identified from the data, we take the union of these estimates to comprise the identified set for the treatment effect of interest (Kamat, Norris and Pecenco, 2023).

We estimate the model using data from the three biggest counties in Ohio, which encompass the cities of Cleveland, Columbus, and Cincinnati. These felony and misdemeanor courts are broadly representative of the American criminal justice system during our study period, reaching from the early 1990s to 2016. Key for our purposes, defendants are randomly assigned to judges, who are responsible for conducting trials and approving any plea deals. We use our method to study the effect of conviction and incarceration—relative to each other and to dismissal—on the number of new charges and convictions in the 5 years following the focal case filing date.

2SLS suggests that incarceration reduces the number of future offenses by 0.362 over the following five years, a 22% reduction relative to the mean, with statistically significant impacts for both felony and misdemeanor incarceration. It also suggests that conviction on more serious felony charges leads to increases in future charges of 0.519 (a 33% increase).

To explore the validity of the 2SLS estimates, we first use our model to estimate treatment effects stripped of bias resulting from exclusion and monotonicity violations. Averaging across felony and misdemeanor courts, we find that the 2SLS estimate using incarceration as the treatment accurately reflects a causal effect; the exclusion and monotonicity terms are small and statistically insignificant. Our methodology further clarifies the 2SLS estimate identifies the effect of incarceration relative to conviction—nearly all of the weight is on compliers who are moved from conviction to incarceration by the instruments. In contrast, the 2SLS estimate for conviction (whether or not the defendant was also incarcerated) is plagued by important violations of exclusion and does not represent the effect of conviction on future crime. Multiple-treatment 2SLS models, which are the most commonly used robustness checks in these settings,

[2] Under latent monotonicity, all compliers between judges with the same conviction propensity are moved between incarceration and conviction without incarceration, and so this method returns valid estimates of the effect of incarceration relative to only conviction when exactly conditioning on this propensity.

do not correct this bias.

Examining the estimates in more detail, we find that the treatment effects differ dramatically between misdemeanor and felony courts. For felony defendants, we find little evidence that conviction affects criminal behavior due to judges largely agreeing on conviction decisions, but we do find that incarceration dramatically lowers future charges and future convictions, by $[-0.379, -0.345]$ and $[-.364, -0.323]$ respectively.[3] Consistent with other recent work (Norris, Pecenco and Weaver, 2021; Rose and Shem-Tov, 2021), the effect of incarceration on crime in each year closely tracks the effects on the number of days spent incarcerated in that year. We conclude that these results are mostly driven by incapacitation effects.

In contrast, we see no effect of misdemeanor incarceration, potentially because sentences are short and hence there is little scope for incapacitation. Conviction, however, increases recidivism by a considerable amount, particularly for defendants without a prior felony conviction. For this group, a conviction causes an additional 0.165 to 0.817 charges over the subsequent five years. The effects are even larger on the number of offenses a defendant is *convicted* of over the next five years, consistent with future police officers, prosecutors and judges being less lenient towards individuals with longer criminal records.

Motivated by the 2SLS-weighted effects, we next use our method to directly estimate the effects of possible policies. We consider reforms that would marginally increase leniency in either the conviction dimension (such as a higher evidentiary standard, or a policy of not prosecuting marginal cases) or the sentencing dimension (such as a change to structured sentencing guidelines to make them more lenient, or more consideration of mitigating factors). Policies increasing leniency in the conviction dimension induce changes in both conviction and incarceration rates and, consequently, require knowledge of treatment effects along both margins. We find that greater misdemeanor conviction leniency would decrease future charges and convictions at essentially no cost. Interestingly, the benefits are particularly large for defendants charged with more serious offenses who would also be incarcerated if convicted.[4] Not all reforms are crime-reducing though—leniency in sentencing would increase future crime in felony courts through decreased incapacitation.[5]

Our analysis relates to several literatures. We first contribute to the extensive body of work using examiner assignment designs. Numerous papers have acknowledged that examiners often choose between more than two options and that this is a threat to causal identification, including in the specific setting of crime we focus on (Bhuller et al., 2020; Norris, Pecenco and Weaver, 2021; Mueller-Smith, 2015). In these cases, they note that 2SLS estimates may be interpreted as causal effects only under strong homogeneity assumptions, such as constant treatment effects across individuals.[6] Our analysis, in contrast, shows how to estimate treat-

[3] Throughout the paper we use square brackets to denote bounds and parentheses for 95% CIs.

[4] We also consider larger policy changes that would eliminate either conviction or incarceration, and find that there might be even larger benefits from policy reforms that target non-marginal defendants.

[5] Incapacitation is an expensive way to reduce crime. We calculate that each averted crime costs between \$55,000 and \$105,000 in prison costs alone, even before accounting for other social impacts.

[6] Bhuller and Sigstad (2022) provides a high-level condition on the compliance groups that restricts two-way flows and under which 2SLS delivers a positively-weighted average of heterogeneous treatment effects. However, this condition is not usually satisfied under standard models of judge decision-making, including the ones we consider in this paper.

ment and policy effects while allowing for heterogeneous treatment effects and flexible choice behavior by judges.

Furthermore, our method provides a blueprint for estimating the effects of multiple treatments in examiner designs that is more general than existing work. As we discuss in Section 5.4, one popular approach has been single-index models of choice (Heckman, Urzua and Vytlacil, 2006; Rivera, 2023). We show that the single-index model is a special case of latent monotonicity, and therefore the bounds that arise from our method will necessarily include the estimates from a single-index approach. Our model is also weaker than other multiple-treatment generalizations applicable to our setup, such as ordered monotonicity (Angrist and Imbens, 1995) and unordered monotonicity (Heckman and Pinto, 2018). In particular, our model allows for what would be traditionally considered defiers—defendants who move out of incarceration when assigned to a more severe judge—as well as two-way flows into and out of conviction. Furthermore, we show that both ordered and unordered monotonicity converge to the single-index assumptions as the number of judges with varying treatment propensities grows, suggesting that our bounds will encompass the estimates from a wide variety of choice models.

Our method can also be seen as more general than recent work that has sought to achieve identification in examiner assignment designs through the use of separability conditions (Humphries et al., 2023, hereafter HOSSD). This approach assumes the existence of additional regressors that shift judge behavior homogeneously in a latent variable space, and hence can be used to point-identify the first stage. Although our method does not require these regressors, if there was a setting where the researcher believed these assumptions were credible, they could be used along with the judge assignment as a source of additional variation. We explore this possibility in Appendix A7, where we estimate their model along with a partially-identified version that differs only in not relying on separability for identification. The estimates arising from the HOSSD method typically lie outside the semiparametric bounds, suggesting that relying on covariates for identification is not innocuous and can substantively affect the results.

In developing these results, our work contributes more generally to the instrumental variable analysis of treatment effects in settings with multiple treatments. Previous work has focused on parameterizations of the selection model that result in point identification (e.g., Hull, 2020; Kline and Walters, 2016). However, an implication of our flexible monotonicity assumption is that point-identifying restrictions for our selection model do not naturally arise. We instead allow the identification of treatment effects under a more flexible, partially identified selection model. Our approach is also distinct from alternatives used in the literature to identify related parameters, which exploit continuous variation in the instrument (e.g., Heckman, Urzua and Vytlacil, 2008; Lee and Salanié, 2018), multiple instruments (Mogstad, Torgovitsky and Walters, 2021; Mountjoy, 2022), or additional data on fallback options (Kirkeboen, Leuven and Mogstad, 2016).

Our approach is most closely related to that of Brinch, Mogstad and Wiswall (2017), which also uses a marginal treatment effect (MTE) framework. They show in a binary treatment setup how simple parametric restrictions on the MTEs allow us to use linear regressions to estimate treatment effect parameters when discrete variation in the instrument does not nonparametrically identify the MTEs—see also the more general framework in Mogstad, Santos and Torgovitsky (2018). However, relative to the binary treatment setup, a key distinction in the multiple treatment case is that the selection model is not generally point identified. Our approach, therefore, shows how a semiparametric assumption on our selection model ensures that we can generally continue to employ several linear regressions to partially identify the parameters of interest. In this sense, our arguments exploit those from Kamat, Norris and Pecenco (2023), which shows that such an insight applies in a general class of selection models.

Finally, we contribute to a rapidly expanding literature studying the effects of incarceration and conviction. Aside from HOSSD and Huttunen, Kaila and Nix (2021), this work has focused on estimating these effects in isolation. In felony courts, incarceration typically leads to reductions in future crime,[7] but prosecution and conviction for misdemeanors and other minor crimes tends to increase recidivism (Agan, Doleac and Harvey, 2022; Mueller-Smith and Schnepel, 2021). It has been unclear whether these disparate findings are caused by differences in the treatment (conviction versus incarceration), type of offense (misdemeanor or felony), research design, or geographic location. We find these results appear to generalize more broadly by replicating the qualitative pattern of effects by treatment and offense severity found across these different studies using a single unified method in a single location.

In addition, our setting and method allow a more comprehensive examination of the misdemeanor criminal justice system than has been previously possible. Although over 11 million people are admitted to jails each year (Zeng, 2020), there are no causal estimates of the effect of post-trial misdemeanor incarceration. Our finding of precisely estimated but statistically insignificant effects stands in contrast to existing work on pre-trial detention, which typically results in higher rates of reoffending (Gupta, Hansman and Frenchman, 2016; Dobbie, Goldin and Yang, 2018; Heaton, Mayson and Stevenson, 2017). One possible reason for this discrepancy is that pre-trial detention also increases conviction rates; consistent with this as the key causal channel, we find that receiving a conviction increases reoffending.

2 Background

2.1 Setting

This study uses data from the courts in Ohio's three largest counties: Franklin County (containing Columbus), Cuyahoga County (containing Cleveland), and Hamilton County (containing Cincinnati). The state is broadly representative of the criminal justice system in the United States in terms of both incarceration and recidivism rates.

Our analysis is based on administrative records collected from the online court information systems. Court records are available starting in the early 1990s (exact date depending on the

[7] For example, some find decreased crime (Rose and Shem-Tov, 2021; Kuziemko, 2012; Norris, Pecenco and Weaver, 2021; Huttunen, Kaila and Nix, 2021; Bhuller et al., 2020), mixed results (Green and Winik, 2010; Estelle and Phillips, 2018; Loeffler, 2013; Harding et al., 2017), and increased crime (Mueller-Smith, 2015).

county) and contain the full case history, including charges, arraignment date, sentencing date and decisions (punishment type and sentence length), and defendant characteristics (name, date of birth, sex, race, and home address). We also use the records to measure future criminal charges and convictions, linking between defendants in different cases using date of birth and name while allowing for slight spelling differences via a fuzzy match. See Norris, Pecenco and Weaver (2021) for details.

Table 1 summarizes the analysis sample. The first column shows that the defendants are disproportionately male (77%), have an average age of 32, and while the broader population of the counties is mostly white, the majority of defendants (61%) are black. Drug and property crimes are the most common offenses (29% of cases each) with fewer for violent (19%), family (14%), and sex (5%) offenses. Defendants have committed an average of 2.2 prior offenses, although the distribution is heavily right-skewed and the median defendant has no past charges.

A key advantage of our setting is the ability to study both misdemeanor and felony cases, which are handled in different courts in each county (Municipal and Common Pleas, respectively).[8,9] Misdemeanors are relatively minor offenses with incarceration sentences no longer than one year; typical examples include soliciting, theft worth less than $1,000 and assault without a deadly weapon. Felonies are more serious and include robbery, theft worth more than $1,000, and assault with a deadly weapon. Felony offenses come with much stiffer criminal penalties. Panel B of Table 1 shows the treatment shares by court; 29% of felony defendants are incarcerated, with a median sentence conditional on incarceration of 615 days. Just 10% of misdemeanor defendants are incarcerated, with a median sentence of only 38 days. Furthermore, conviction rates are higher in felony courts (87 versus 53%) and a felony criminal record is typically viewed as more serious, potentially affecting labor market opportunities going forward (Agan et al., 2023).

2.2 Using judges as instruments

Importantly for the design of this study, Ohio law mandates that most criminal cases are randomly assigned to judges. The random assignment is carried out by a computer program and done separately by court after charges are filed.[10] Defendants with ongoing cases or who are on probation are excluded from this randomization, although this amounts to a minority of cases. We drop all non-randomly assigned cases from our sample, and conduct our analysis using the identity of the first-assigned judge to account for any issues arising from the approximately 5% of cases who are transferred between judges due to workload and scheduling issues. To restrict comparisons between the set of judges available at any given

[8] We focus on misdemeanor cases of 1st to 4th degree to focus on cases commonly considered criminal. Municipal courts also handle traffic cases and minor misdemeanors, which are very low-level offenses such as noise ordinance complaints.

[9] While a charge can be reduced from a felony to a misdemeanor throughout the court process, we classify each case by where it is originally filed and refer to Common Pleas cases as felony cases whether or not the judge reduces the severity of the charge.

[10] This means that a decision on whether the defendant will be held pre-trial has already been made at the time of judge randomization. These decisions are made by a separate bail judge, and although they have important effects on defendants (Dobbie, Goldin and Yang, 2018) analyzing them is outside the scope of this study.

time, throughout our analysis we include court-year fixed effects. Consistent with random assignment, the last two columns in Table 1 show that the judges' incarceration and conviction severity are uncorrelated with observable characteristics of defendants (p=0.76 and p=0.35, respectively) conditional on stratifying fixed effects.

In criminal cases, courts make two sequential decisions: first, whether to find the defendant guilty, and second, if they are found guilty, what the sentence will be. The judge has an important influence on both of these decisions.

After a defendant has been arraigned, the judge guides the proceedings towards a trial. She oversees the pre-trial procedure, which includes setting the schedule (which can affect lawyers' preparedness) and determining which evidence will be admissible. These initial decisions can dramatically affect the likelihood of success at trial, and cases are sometimes dropped by prosecutors after a string of unfavorable pre-trial decisions.

Defendants in Ohio have a right to a trial by jury. If they choose to exercise this right, the judge oversees the jury selection process and then provides instructions to the jury, which votes on whether to find the defendant guilty. If the defendant opts against a jury trial, the judge makes the ruling herself. Whether or not it is a jury trial, however, the decision on guilt is supposed to be made without regard to the possible sentence: jurors are not even told the range of possible punishments in an attempt to stop this information from affecting their conviction decision. As we discuss in Section 4, we mirror this institutional structure in our identification strategy.

For defendants who have been found guilty, the judge determines the sentence from an offense-specific allowable range that often includes the possibility of either probation or an incarceration sentence.[11] The determinants of the sentence are different from the factors that affect conviction and include whether there were mitigating or aggravating features of the offense, as well as details of the defendant's criminal and personal history. To concentrate attention on the legally relevant facts, sentencing usually occurs only after a presentencing investigation (PSI) that contains this information (as well as victim input) has been completed.[12] Importantly, judges' sentencing decisions are *not allowed* to reflect their prior on the defendant's guilt.

One important institutional factor in Ohio courts is the prevalence of plea deals. Approximately 88% of felony cases end in this way, with a guilty plea from the defendant before trial and a joint recommendation from the prosecutor and defense attorney on the sentence.[13] The judge then proceeds with sentencing (usually after receiving a PSI), taking the recommendation into consideration. While a large literature has studied plea deals (Landes, 1971; Priest and Klein, 1984; Bebchuk, 1984; Silveira, 2017), to our knowledge there has not been an anal-

[11] One other way that judges affect sentences is through changing the degree of offense the defendant is convicted of; for example, by determining the value of a theft was less than $7,500, which reduces the offense level from a third degree felony to a fourth degree felony and decreases the maximum prison time.

[12] For felony cases, until 2016 a PSI was required before imposing probation or a sentence shorter than six months. Since one of these punishments is usually an option for most offenses (and because most judges like to have a PSI even when determining the length of incarceration), PSIs are completed in most cases. They are not required in misdemeanor cases, but judges often request them anyway.

[13] Data on plea deal prevalence comes from the felony court in Cuyahoga, which has slightly higher conviction rates than the other courts.

ysis of how plea bargaining might affect or invalidate choice models based on the institutional details of a non-plea system. We study this problem in Appendix A1 and show that while the introduction of plea bargaining might change the treatment that a particular defendant is assigned to, in a full-information Nash bargaining game, the same two-stage process can be used to describe the choice model with and without plea bargaining. We return to this issue in Section 4, when we discuss concrete ways that plea deals could undermine our choice model.

3 IV interpretation and results

In this section we present simple 2SLS estimates of the effects of incarceration and conviction on outcomes. We then discuss the possible challenges to identification and interpretation that arise from the multiple margins of treatment in our setting.

3.1 IV estimates

The identity of the assigned judge is commonly used as an instrument for treatments such as incarceration or conviction. This research design is typically justified by the monotonicity and exclusion conditions in Imbens and Angrist (1994), which assumes that moving from a lower to a higher treatment propensity judge makes all defendants weakly more likely to receive the focal treatment and does not move any defendants between the non-focal treatments. In this section, we present 2SLS estimates based on this research design.

Given our focus on the conviction and incarceration margins, we aggregate the set of judicial decisions to be $D = \{n, c, p\}$, where n denotes no conviction, c denotes conviction without incarceration, and p denotes conviction with incarceration.[14] We analyze treatments T that equal $D^p = \mathbb{1}[D{=}p]$, indicating an individual is incarcerated, or $D^{cp} = \mathbb{1}[D \in \{c, p\}]$, indicating any form of criminal conviction. Individuals are assigned to a judge $Z \in \mathcal{Z} = \{z_0, z_1, \ldots, z_J\}$.

For each of these treatments, we use the following 2SLS specification:

$$Y_i = \beta^{\text{2SLS}} T_i + \mu_x + \varepsilon_i \tag{1}$$

$$T_i = \sum_{j=1}^{J} \alpha_j \mathbb{1}[Z_i{=}z_j] + \phi_x + e_i \tag{2}$$

where Y_i is the outcome and μ_x and ϕ_x are court-year fixed effects as required by the design.[15]

Table 2 reports the estimated effects of incarceration (Panel A) and criminal conviction (Panel B) on cumulative number of charges over the following five year period. If the Imbens and Angrist (1994) assumptions are satisfied, column (1) in Panel A shows that incarceration reduces the number of future charges by 0.36. Relative to the mean of 1.6 charges over the next 5 years, this is a consequential 23% decrease. The other columns show additional

[14] We pick p for the mnemonic with "prison," although not all incarcerated defendants technically go to prison, which is for sentences longer than a year. Those who are incarcerated on shorter sentences go to jail.

[15] Judge instruments are sometimes constructed as leave-one-out averages to ameliorate potential finite sample bias issues. Because there are many observations per judge, this is empirically unimportant in our setting.

heterogeneity to contextualize future results. Columns (2) and (3) show the results separately in the felony and misdemeanor courts. Incarceration significantly reduces future charges by 0.48 for defendants in the felony court, while there is a marginally statistically significant ($p < 0.10$) reduction of 0.13 charges for misdemeanor defendants. Although there has been much recent interest in the potential impacts of criminal justice sanctions for individuals who have not previously been convicted of a felony, columns (4) and (5) show similar impacts for this population compared to the full population in both courts.

Turning towards the estimated effects of conviction in Panel B, column (1) shows little average effect, although this result masks important heterogeneity across offense types. Column (2) suggests that conviction in felony courts is highly criminogenic, increasing future charges by 0.52, while column (3) finds a statistically insignificant and smaller decrease in future charges arising from a misdemeanor conviction. Columns (4) and (5) report similar results for the sample without a previous felony conviction.

The validity of the arguments in Imbens and Angrist (1994) and its extensions is based on the assumption of a single treatment. In our context, this translates to the requirement that the assignment to a higher treatment propensity judge makes all defendants weakly more likely to be assigned that treatment (monotonicity) and doesn't move any defendants between the non-focal treatments. The judges' ability to affect both conviction and incarceration raises concerns about the causal interpretation of the above conclusions. A common attempt (e.g., Mueller-Smith, 2015; Bhuller et al., 2020; Norris, Pecenco and Weaver, 2021) to account for this concern is to run an augmented 2SLS specification where the second stage includes both treatments, and where each treatment is instrumented for with the judge assignment as follows:

$$Y_i = \beta_{incar}^{\text{2SLS}^*} \mathbb{1} D_i^p + \beta_{convic}^{\text{2SLS}^*} \mathbb{1} D_i^{cp} + \mu_x + \varepsilon_i \tag{3}$$

$$D_i^p = \sum_{j=1}^{J} \alpha_j^1 \mathbb{1}[Z_i = z_j] + \phi_x^1 + e_i^1 \tag{4}$$

$$D_i^{cp} = \sum_{j=1}^{J} \alpha_j^2 \mathbb{1}[Z_i = z_j] + \phi_x^2 + e_i^2. \tag{5}$$

Panel C in Table 2 reports estimates of the coefficients $\beta_{incar}^{\text{2SLS}^*}$ and $\beta_{convic}^{\text{2SLS}^*}$. While the results for the impacts of incarceration and overall effects of conviction are unchanged, column (2) shows a somewhat weaker effect of felony conviction when controlling for incarceration. These broadly similar results may support the empirical conclusions of the single-treatment regressions, although previous work has noted the challenges in interpreting this multiple-treatment specification. In the next section, we develop a framework for interpreting these estimates further.

3.2 When is 2SLS interpretable?

To provide a basis for our analysis and to highlight how the presence of multiple treatments makes the 2SLS assumptions more onerous, in this section we provide a simple decomposition of the 2SLS estimand into its constituent parts: the target parameter, exclusion violations, and monotonicity violations. To our knowledge, this result is novel in accommodating more than two judges, which is key to decomposing the 2SLS estimates in practice. In Section 6.3 we apply this result in our setting.

For most of this section, we focus on the single-treatment 2SLS estimand in (1) when one instruments for incarceration, but a similar analysis applies if one instead instruments for conviction. We discuss the multiple-treatment case at the end of the section and provide the analogous decomposition in Appendix A3.

We represent the observed decision and future criminal behavior using standard potential outcomes notation, so $D(z)$ denotes the potential decision had the individual been assigned to judge z, and $Y(d)$ denotes their potential outcome under treatment d. We take the judge as an exogenous instrument conditional on covariates X, which in our setting are court-year indicators that control the set of judges available for random assignment to the case:

Assumption E *(Exogeneity) Z is jointly independent of $\big(Y(n), Y(c), Y(p), D(z_1), \ldots, D(z_J)\big)$ conditional on X.*

Our notation also implicitly assumes that the instruments are excludable conditional on the three treatments: judges can affect outcomes only through the treatments of interest, rather than through other channels such as variation in sentence length for incarceration always-takers or through differences in probation conditions for conviction always-takers.

The building block of our decomposition is the compliance group, which is defined by the vector of treatment responses across judges, $(D(z) : z \in \mathcal{Z})$. For any two treatments s and t, individuals in each compliance group are on average either induced from s into t by the instruments, or from t into s. As a result, we define $\Delta_{s \to t}$ as the $s \to t$ effect for those compliance groups who are on average induced between s and t by the instruments, with the corresponding weights $\phi_{s \to t}$ for each compliance group determined by how much they contribute to the 2SLS estimand. This leads to the following decomposition:

Proposition 1 *Under Assumption E, the 2SLS estimand in (1) with incarceration as the treatment can be decomposed as*

$$\beta^{2SLS}_{incar} = \underbrace{\phi_{c\to p}\Delta_{c\to p} + \phi_{n\to p}\Delta_{n\to p}}_{\textit{effect of incar. rel. to alternatives}} + \underbrace{\phi_{n\to c}\Delta_{n\to c} + \phi_{c\to n}\Delta_{c\to n}}_{\textit{exclusion violations}} + \underbrace{\phi_{p\to c}\Delta_{p\to c} + \phi_{p\to n}\Delta_{p\to n}}_{\textit{mono. violations}} \qquad (6)$$

where $\phi_{s\to t} \geq 0$ for all $s, t \in \{n, c, p\}$ and

$$\phi_{c\to p} + \phi_{n\to p} - \phi_{p\to c} - \phi_{p\to n} = 1$$

Proof and definition of ϕ: see Appendix A2.

This proposition reveals that $\beta^{\text{2SLS}}_{incar}$ can be viewed as containing three components: the weighted effect of being in treatment p relative to the non-carceral treatments n and c, the weighted effect of being in the non-carceral treatments relative to p, and the weighted effect of being moved between the two non-carceral treatments.[16] Interpretation of $\beta^{\text{2SLS}}_{incar}$ as a treatment effect of incarceration relative to alternatives therefore requires eliminating the latter two components of the decomposition.

The most natural way to remove the exclusion violations is through an additional assumption of *binary exclusion*, or that for all individuals for whom there exist instrument values z, z' such that if $D(z)=n, D(z')=c$, then $Y(n) = Y(c)$. This condition can be satisfied in two ways: either there are no individuals who are moved between n and c by changes in instrument assignment (so $\phi_{c\to n} = \phi_{n\to c} = 0$), or receiving the treatment n rather than c does not affect the outcomes of the compliers between n and c (so $\Delta_{c\to n} = \Delta_{n\to c} = 0$). Researchers may be unsure that either of these assumptions holds; the first because judges differ dramatically in their conviction rates, and the second because of evidence of the effect of convictions on outcomes in other settings (Pager, 2003; Agan and Starr, 2018; Agan, Doleac and Harvey, 2022).

Similarly, removing the monotonicity violations requires restrictive assumptions on judge behavior. The traditional binary-treatment monotonicity assumption requires that no individual is moved from incarceration into one of the other treatments when she is assigned to a higher-incarceration-propensity judge; this implies that $\phi_{p\to n} = \phi_{p\to c} = 0$. However, this assumption is at odds with the different standards for conviction and incarceration in our setting, as highlighted in Section 2.2. Since incarcerated defendants might face the fallback treatment of either conviction (if the evidence they committed the crime was strong but the sentencing guidelines did not require incarceration) or not guilty (if the evidence they committed the crime was middling but the sentencing guidelines made incarceration mandatory if convicted), this suggests that judges may often differ in their fallback options, violating monotonicity.

These pitfalls of single-treatment 2SLS, unfortunately, are not generally overcome by instrumenting for both treatments simultaneously. In Appendix A3, we provide a detailed analysis of the 2SLS specification in (3), and show the estimands can be decomposed into compliance group-weighted effects across different treatment comparisons as in (6). These effects are a combination of the parameters from the single-treatment case, and neither can typically be interpreted as a positively-weighted treatment effect.

Contemporaneous work has sought to clarify what additional assumptions can be imposed for 2SLS to identify causal effects in multiple-treatment models. In particular, Bhuller and Sigstad (2022) shows that two high-level assumptions—average causal monotonicity and no cross effects—are sufficient to generate properly-weighted treatment effects. However, these assumptions impose substantial uniformity in how judges make decisions, and imply that the relative effect of changing c and p judge propensities on defendants' likelihood of receiving treatment p must be the same for all compliers.[17] More intuitively, in Appendix A3 we show

[16] Interpretation of $\beta^{\text{2SLS}}_{convic}$ is analogous, but with different terms labelled as exclusion and monotonicity violations.

[17] An alternative (Proposition B.8) is to adopt a stringent Imbens-Angrist monotonicity condition for each judge and treatment, as well as the empirically testable assumption that the judge propensities for c and p are linear in expectation. The monotonicity assumption, however, rules out substitution patterns likely to exist in examiner contexts such as two-way flows in and out of c.

how even when the choice model is such that judge-pair comparisons could be used to isolate margin-specific causal effects, 2SLS fails to isolate these comparisons.

In both the single- and multiple-treatment cases, these interpretational issues are lessened if the treatment effects are constant across individuals.[18] However, constant effects has testable implications, one of which is that the model will not reject overidentifying restrictions. In Table 2 we report the results of J tests of overidentification, and reject with p-values of 0 in each type of court as well as overall.

With treatment effect homogeneity rejected and existing choice assumptions inappropriate for our setting, a credible analysis requires two things: a set of assumptions on the instruments that arises naturally from judge behavior in this context, and an estimation methodology that credibly isolates the instrument-induced changes. The subsequent sections do this in turn.

4 A more flexible choice model

Our goal is to provide a choice model that matches the institutional environment and accommodates the key ways that judge assignment might affect defendants' treatment assignment. In particular, certain judges might be more likely to find defendants guilty, perhaps because they are more likely to admit evidence against defendants. Other judges might be more likely to incarcerate defendants because they have a less lenient understanding of what the sentencing guidelines require upon conviction. Finally, some judges might be relatively likely to find defendants guilty *and* to sentence defendants to prison.

The last pattern is particularly challenging for identification because it raises the possibility of two-way flows: assignment to judge z rather than z' might move some defendants from n (not guilty) to c (convicted but not incarcerated), and different defendants from c to p (prison). Most existing work (e.g. Heckman and Pinto, 2018) disallows two-way flows in order to point-identify the share of each complier group.

To address these issues, our choice model matches the legal requirement that judges' conviction decisions are separate and independent of their sentencing decisions. The index representation of this model is originally due to Arteaga (2021); we provide a fuller characterization that links it to an equivalent potential outcomes form. This characterization facilitates a new result on the flexibility of latent monotonicity over other discrete choice models in examiner settings. We then introduce a new way to estimate all treatment effects—such as conviction relative to dismissal—rather than just using 2SLS to estimate the effect of incarceration relative to conviction as in Arteaga (2021). The cost of this flexibility—and in particular, the accommodation of different types of two-way flows—is that the first stage is not point-identified, and as a result the estimated treatment effects are bounds.

4.1 Latent monotonicity

As we discuss in Section 2.2, the court's decision-making process proceeds in two steps: they first assess whether the defendant is guilty, and if they are guilty, decide the punishment.

[18] In the single-treatment case, one still needs to assume binary exclusion.

These separate decisions depend on different criteria and sets of facts, and are supposed to be made separately.

In line with this system, we treat the court as making two sequential decisions: first, whether to convict, and second, what sentence to impose. We assume that judges behave monotonically with respect to conviction, and *would* behave monotonically with respect to incarceration during the sentencing phase. More precisely, let $D_p^*(z)$ denote an indicator for whether an individual is potentially incarcerated by judge z in the hypothetical case where they were already convicted. This means that the potential treatment for whether the individual is incarcerated is given by $D_p(z) = D_{cp}(z)D_p^*(z)$. This allows the following monotonicity condition:

Assumption LM *(Latent Monotonicity) For each $z, z' \in \mathcal{Z}$, we have*

$$\begin{aligned} D_{cp}(z) \geq D_{cp}(z') \quad &or \quad D_{cp}(z) \leq D_{cp}(z') \ , \\ D_p^*(z) \geq D_p^*(z') \quad &or \quad D_p^*(z) \leq D_p^*(z') \ . \end{aligned}$$

Heuristically, this condition imposes that a judge's incarceration decision would not hinge on whether she thinks the appropriate guilt standard has been met, in line with the legal separation between these standards. This rules out certain types of forward-looking behavior by both the judge and the broader legal system. First, the monotonicity condition on $D_{cp}(z)$ implies that if two judges have the same conviction rate, they must convict the same individuals. This rules out situations where one of the judges realizes that she would incarcerate a particular defendant if she convicted him and, in response, decides not to convict him in the first place. If the second judge decides to convict (treatment c or p), this violates LM.

Similarly, latent monotonicity rules out certain types of behavior by prosecutors. Suppose that judge z has a lower conviction rate (treatment c or p) than judge z', so it must be that $D_{cp}(z) \leq D_{cp}(z')$ under LM. Imagine that conditional on conviction, a particular defendant would be incarcerated by judge z but only convicted (treatment c) by judge z'. If the prosecutors decide that only a carceral sentence would be worth their time and drop the case under z (but not under z'), then $D(z) = p$, $D(z') = n$ for this defendant and LM is violated.

Another possible threat to our choice model comes from plea bargaining. As discussed in Section 2.2, most cases in Ohio end in a plea bargain, not in a trial. To understand the possible implications of plea bargaining, in Appendix A1 we build a simple model of a prosecutor and defense attorney who Nash bargain over the possible sentence. We find that while plea bargaining can change the outcome of particular cases, if LM is an appropriate model of the choice process without plea bargaining it will continue to be appropriate after the introduction of plea bargaining.

The advantage of latent monotonicity is that by mirroring the institutional environment, it allows for more choice patterns than other available alternatives—and precisely those required by the setting. Table A7 shows the possible compliance patterns under LM as well as for ordered monotonicity (Angrist and Imbens, 1995), unordered monotonicity (Heckman and Pinto, 2018), and single-index choice models (Heckman and Vytlacil, 2005; Bhuller and Sigstad, 2022; Rivera, 2023). In contrast to these other assumptions, LM allows for two-way flows in both conviction and incarceration as well as for defendants to be marginal between all three treatments. We discuss these issues further in Appendix A4, and provide one additional result that shows that the choice patterns under ordered and unordered monotonicity can become increasingly restricted as the number of judges increases; LM does not.

It will be convenient to work with a slightly different form of LM for estimation. In the following proposition, we show how latent monotonicity can be recast as an index assumption.

Proposition 2 *Given Assumption E, Assumption LM is equivalent to*

$$D(z) = \begin{cases} n & \textit{if} \;\; U_1 > g_1(z) \;, \\ c & \textit{if} \;\; U_1 \leq g_1(z) \;,\; U_2 > g_2(z) \;, \\ p & \textit{if} \;\; U_1 \leq g_1(z) \;,\; U_2 \leq g_2(z) \;, \end{cases} \tag{7}$$

where $U_1 \sim U[0,1]$ and $U_2 \sim U[0,1]$, and $(g_1(z), g_2(z)) \in [0,1]^2$ are judge-specific thresholds. Let F denote the cumulative distribution function of (U_1, U_2).
Proof: see Appendix A5.

The above proposition shows that Assumption LM is equivalent to imposing a two-stage threshold crossing equation for each judge's decision for an individual. Judges first decide whether to convict an individual or not, and then, conditional on conviction, they decide whether to incarcerate them or not. Panel A of Figure 1 provides a graphical description of the threshold structure in the space of the individual latent variables. U_1 can be interpreted as an individual's resistance to conviction, since judges first convict those with lower values, while U_2 can be interpreted as their resistance to incarceration since judges would first incarcerate those with lower values. Analogously, $g_1(z)$ and $g_2(z)$ can be respectively interpreted as a judge's level of severity along the conviction and incarceration margins, as those with higher values respectively convict and incarcerate more defendants. In the special case where $U_1{=}U_2$, which corresponds to all judges evaluating defendants along only one dimension of criminality, LM is equivalent to the single-index model.[19] However, other distributions of F will, in general, correspond to richer substitution patterns and can capture the different dimensions of criminal cases.

While we favor this monotonicity assumption and the associated threshold model primarily because it shares a number of key features with the institutional context, we can also use our data to assess the model implications. Appendix A6 derives testable implications of both LM and a single-index model on judge-pair Wald estimands over appropriately defined outcome moments and performs these tests. We reject these implications for the single-index model but not for LM, providing support for our choice model.

[19] See Appendix A4 for a formal definition of the single index model.

5 Identification of treatment effects

We develop a novel methodology to apply this choice model to the data. The main identification challenge is that the parameters of the selection equation—and thus, the size and even existence of compliance groups—are not point-identified from the data. In this section we show how we can nonetheless partially identify the first stage and then use polynomial marginal treatment response functions (Brinch, Mogstad and Wiswall, 2017) to partially identify a wide range of parameters of interest, including 2SLS-weighted treatment effects and policy-relevant treatment effects.

5.1 Indeterminacy of the first stage

The selection equations in (7) relate the parameters of the selection model to treatment decisions. They imply that for each judge, the treatment propensities are functions of those parameters:

$$P(D{=}n|Z{=}z, X{=}x) = 1 - g_1(z,x) \ , \tag{8}$$

$$P(D{=}c|Z{=}z, X{=}x) = g_1(z) - F_{U_1,U_2}\big(g_1(z,x), g_2(z,x)\big) \ , \tag{9}$$

$$P(D{=}p|Z{=}z, X{=}x) = F_{U_1,U_2}\big(g_1(z,x), g_2(z,x)\big) \ . \tag{10}$$

where we now add x as an argument to g to emphasize that our first-stage identification is conditional on covariates.

From (8), it is clear that $g_1(z,x)$ is directly identified from the data. However, (9) indicates that $g_2(z,x)$ depends crucially on the unobserved joint distribution of U, F.[20] Different distributions of F correspond to different values of $g_2(z,x)$ and thus different compliance patterns.

We illustrate the indeterminacy of the first stage in Panels B and C of Figure 1. The figure shows the response types between two judges when U is distributed as a normal copula with unknown correlation ρ. We assume that judge z incarcerates 10% of defendants and convicts 10%, while judge z' incarcerates 20% and convicts 60%. These treatment shares result in different compliance patterns for different values of ρ. In Panel B, where $\rho{=}0$, 15% of the population is a $n{\to}p$ compliers and 5% is a $p{\to}c$ complier. There are no $c{\to}p$ compliers. However, when $\rho{=}0.8$ (Panel C), there are no $p{\to}c$ compliers and instead 2.9% of the population is a $c{\to}p$ complier. Therefore, although for each state of the world, changing judges from z to z' results in a 10 p.p. net increase in the incarceration rate, the size and existence of the compliance groups differs dramatically across values of ρ.

Our challenge is to continue to learn about treatment effects despite non-identification of the first stage. To make progress on this front, we assume that U is distributed in a parametric family known up to a parameter $\rho \in P$.[21] This substantially simplifies our task, because it

[20] By adding up of the treatment shares, (10) adds no additional information on g.

[21] In our baseline analysis, we assume that F is normal copula, although we also consider U distributed according to a copula based on the logistic distribution (Ali, Mikhail and Haq, 1978).

means that for a given value of ρ we can directly calculate $g(z)$ using (8)-(10) and thus identify the compliance groups between each pair of judges.

5.2 Parameters of interest

Our analysis exploits the fact that the parameters of interest can be expressed as known functions of the same primitives that determine the observed data. We take the primitive of the selection equation for the judge decisions to be the correlation ρ between the components of U, noting that ρ determines the judge thresholds g. Following Heckman and Vytlacil (1999, 2005), we take the primitives with respect to the outcomes to be the marginal treatment response (MTR) functions

$$m_d(u_1, u_2, x) \equiv E[Y(d)|U_1{=}u_1, U_2{=}u_2, X{=}x]$$

for $d \in \mathcal{D}$. We can then express our parameters of interest in terms of the MTRs and the unknown primitive of the selection equation, ρ. For example, a key object of interest is a version of the 2SLS estimand stripped of the exclusion and monotonicity violations present in (6). Specifically, we take the weighted average of the $n{\to}p$ and $c{\to}p$ effects:

$$\beta_{(n,c)\to p} = \omega_{c\to p}\Delta_{c\to p} + \omega_{n\to p}\Delta_{n\to p} \quad , \quad \omega_{s\to t} = \frac{\phi_{s\to t}}{\phi_{c\to p} + \phi_{n\to p}} \tag{11}$$
$$\Delta_{s\to t} = \sum_{x\in\mathcal{X}} w_x \int [m_t(u_1,u_2,x) - m_s(u_1,u_2,x)] w_{s\to t}(u_1,u_2,x) dF(u_1,u_2)$$

where w_x is the weight on covariate cell x. This weight, as well as the weights on the different compliance groups, $w_{s\to t}$ and $\phi_{s\to t}$, are defined in Appendix A2, where we show that these weights are also functions of ρ.

We also consider how to estimate the effect of particular policy reforms. For example, consider a policy that changes the evidentiary burden such that a judge becomes δ more lenient on the conviction margin. This will move some defendants from p to n, and others from c to n. Using $\mathcal{P}^{\delta}_{s\to t}(z,x)$ to denote the types who are moved from treatment s to t by the policy change,[22] the effect of the policy on outcomes is

$$\underbrace{\int_{u\in\mathcal{P}^{\delta}_{p\to n}(z,x)} [m_n(u_1,u_2,x) - m_p(u_1,u_2,x)]dF(u_1,u_2)}_{\text{dismissal effect for counterfactually incarcerated defendants } (p\to n)} + \underbrace{\int_{u\in\mathcal{P}^{\delta}_{c\to n}(z,x)} [m_n(u_1,u_2,x) - m_c(u_1,u_2,x)]dF(u_1,u_2)}_{\text{dismissal effect for counterfactually convicted defendants } (c\to n)}$$

which is a function of both how many defendants are moved into n from p versus c, as well as the magnitude of the treatment effect for each of these groups. In Section 6.7 we report the effects of various policies that adjust judges' thresholds.

[22] E.g., $\mathcal{P}^{\delta}_{p\to n}(z,x) = [g_1(z,x) - \delta, g_1(z,x)] \times [0, g_2(z,x)]$.

5.3 Regression-based solution to identification problem

Just as the parameters of interest can be expressed as functions of the underlying primitives, so can the observed data. Our challenge is to learn which MTRs are consistent with the data and hence which values of the parameters of interest are also consistent with the data.

To do so, we assume the MTRs are polynomials in u_1 and u_2, where we allow the MTRs to differ across covariate cells:

$$m_d(u_1, u_2, x) = \sum_{k_1=0}^{K_{d1}} \sum_{k_2=0}^{K_{d2}} \alpha_{dk_1k_2x} u_1^{k_1} u_2^{k_2}$$

This allows us to write expected outcomes given judge assignment and treatment as a simple linear function of calculable regressors. In particular, we have that

$$\begin{aligned} E[Y \mid D{=}d, Z{=}z, X{=}x] &= \frac{1}{P_{dxz}} \int\limits_{u \in \mathcal{U}_{dxz}(\rho)} m_d(u_1, u_2, x) dF(u_1, u_2) \\ &= \sum_{k_1=0}^{K_{d1}} \sum_{k_2=0}^{K_{d2}} \alpha_{dk_1k_2x} \underbrace{\int\limits_{u \in \mathcal{U}_{dxz}(\rho)} \frac{u_1^{k_1} u_2^{k_2}}{P_{dxz}} dF(u_1, u_2)}_{\equiv h_{dk_1k_2xz}(\rho)} \end{aligned}$$

where P_{dxz} is the likelihood that judge z in covariate cell x assigns treatment d and $\mathcal{U}_{dxz}(\rho)$ is the rectangular area in the space of unobservables such that a defendant with covariate x and index u would receive treatment d if assigned to judge z.[23]

For each value of ρ, there is a different distribution of U (through F) as well as a different mapping between u and treatment (through $\mathcal{U}_{dxz}(\rho)$). This implies a different relationship between the selection indices u and outcomes, which is summarized by $h(\rho)$. It also suggests that for each value of ρ, this relationship can be recovered by a simple linear regression of outcomes on the calculable covariates h:

$$y_{idxz} = \sum_{k_1=0}^{K_{d1}} \sum_{k_2=0}^{K_{d2}} \alpha_{dk_1k_2x}(\rho) h_{dk_1k_2xz}(\rho) + \varepsilon_{idxz} \tag{12}$$

To implement our model, we allow the MTRs to vary across each of the six courts in our data, with the treatment-specific intercepts additionally varying at the year level. We view this as a middle ground between MTRs that do not vary with X—and thus use cross-court variation in h to identify the selection parameters of the MTRs—and allowing the MTRs to vary flexibly by court-year.[24]

This means that there are $6|D|((K_{d1}+1)(K_{d2}+1)-1)+|D||X|$ parameters, and $|D||X||Z|$ moments. There are therefore many more moments than parameters, and $\alpha(\rho)$ is point-

[23] Specifically, $\mathcal{U}_{nxz}(\rho) = [g_1(z,x), 1] \times [0,1]$, $\mathcal{U}_{cxz}(\rho) = [0, g_1(z,x)] \times [g_2(z,x), 1]$, and $\mathcal{U}_{pxz}(\rho) = [0, g_1(z,x)] \times [0, g_2(z,x)]$.

[24] The assumption of some separability on the MTRs is common practice in the applied MTE literature as this expands the region of identification of the outcome functions (Cornelissen et al., 2016). However, we emphasize that we do not use separability to identify the first stage.

identified for each value of ρ if the standard regression rank condition is satisfied.

As discussed in the previous section, our objects of interest are functions of the MTRs and ρ. Letting $\theta(\alpha, \rho)$ represent a generic parameter of interest (e.g., $\beta_{(n,c)\to p}$ in (11)), for any value of ρ we can calculate this parameter as $\theta\big(\alpha(\rho), \rho\big)$. However, since ρ is not identified from the data, there are a range of possible values of θ that will be consistent with the data. To summarize these values, we take the union of values of θ across different plausible values of ρ as our identified set (Kamat, Norris and Pecenco, 2023). Formally, the identified set is

$$\Theta = \Big\{\theta_0 : \theta_0 = \theta\big(\alpha(\rho), \rho\big) \text{ for some } \rho \in P\Big\}.$$

Because h is a nonlinear function of ρ that must be calculated using numerical quadrature, we estimate the identified set by calculating h for values of ρ over a finite-size grid. For each ρ we then estimate (12) and use these estimates to calculate θ. We then take the smallest and largest values of θ across this grid as the lower and upper bounds on the parameter of interest.

5.4 Connection to other approaches

A number of recent papers have considered identification of multiple treatment effects in settings with either additional data on fallback options (Kirkeboen, Leuven and Mogstad, 2016) or instruments that vary in multiple dimensions (Mountjoy, 2022). We lack data on outside options and have only a single-dimensional discrete instrument—the assigned judge—so we cannot directly apply either of these approaches.

The Arteaga (2021) study of the effect of parental incarceration on child outcomes also considers the threshold selection model in (7). We make several additional contributions. First, we clarify which assumptions on potential outcomes give rise to this model and connect it to other popular choice models. Second, because of data limitations, Arteaga (2021) uses the model only to motivate 2SLS regressions of outcomes on instrumented incarceration among the sample of convicted defendants, controlling for judges' conviction propensity. In contrast, we show how the same choice model can be used to estimate a variety of additional treatment effects, including 2SLS-weighted effects of conviction relative to both incarceration and case dismissal, as well as other policy-relevant treatment effects.

Our approach shares some similarities with Humphries et al. (2023), which also considers a setting with three treatments and examiner instruments. Their idea is that one might be able to use tools from the industrial organization literature to transform the single-dimensional judge assignment into a multi-dimensional instrument and then apply Mountjoy (2022) to point-identify the treatment effects. However, as we show in Appendix A7, their approach relies on a strong separability condition where a special class of regressors varies within judge and affects decisions for all judges equally. Since it is unclear what regressors might satisfy this non-standard and difficult-to-test assumption, it's not clear why imposing homogeneity on judge decision-making is necessarily an improvement over assuming homogeneity in treatment effects, the typical approach used with multiple-treatment 2SLS. It also means that in the usual best-case scenario of unconditionally randomly assigned judges and no additional covariates—

or of judge effects that are nonseparable in the covariates—their model is not identified.

When their identification assumptions are satisfied, separability allows point-identification of the first stage, and as a result point-identification of treatment effects. These estimated treatment effects should lie inside the bounds produced by our approach, resulting in more informative conclusions for the researcher. In Appendix A7, we use our data to explore this testable implication, and, as a result, the credibility of the separability assumption. We first estimate a version of their baseline model as faithfully as possible, using the same choice model, functional forms, identifying regressors, outcome model, and estimands. We then show how one can adopt our first stage identification strategy to their model and estimate a partially-identified specification that differs only in not using the additional regressors for identification. The baseline estimates fall *outside* the bounds in many subsamples, indicating that the Humphries et al. (2023) identifying restrictions on the covariates are not satisfied in our setting and can result in substantively different conclusions.

Finally, our approach is also closely related to the single-index model (Rivera, 2023), which is a special case of our model when $\rho = 1$. Our bounds therefore encompass the single-index estimates.

5.5 Implementation details

In our main analysis, we assume that F is a normal copula with correlation ρ, although as a robustness check we also allow U to be distributed in a logistic family. We further assume that $\rho \geq 0$, so that defendants who have unobservable characteristics that make them more likely to be convicted also have unobservable characteristics that make them more likely to be incarcerated. This accounts, for example, for the fact that defendants with more serious criminal records may be more likely to be convicted (because prosecutors are more motivated to obtain a guilty verdict) and that they are more likely to serve an incarceration sentence (because sentences tend to get longer with one's criminal record (Shen et al., 2020)).

As previously discussed, we calculate g at the judge-year level. This means that for each value of ρ, there is a single set of thresholds g that satisfies the treatment shares in (8)-(10). We assume that the MTRs are partially separable, with a court-year-specific intercept and terms in u that vary by court. Since judges work in only one court, this means we exploit only within-court-year variation in judges' decisions to identify the selection terms.

We assume that the c and p MTRs are second-order in each dimension of u, although we impose that the coefficient $\alpha_{d22x} = 0$ for numerical stability. Similarly, the n MTR is assumed to be second-order in u_1 but constant in the u_2 dimension, since all cross-judge variation in u_2 relevant to the n MTR is driven by changes in the u_1 dimension. We calculate the regressors $h(\rho)$ using numerical integration, then estimate $\alpha(\rho)$ and the corresponding object of interest for each $\rho \in \{0, 0.2, ..., 1\}$. We report the upper and lower bounds across ρ. For inference, we estimate the covariance matrix of the estimates over ρ using 200 bootstrap draws, and then use Bei (2023) to construct a confidence interval on the union bound.

We study two main types of parameters. First, we study the 2SLS-weighted net complier effects discussed in Section 5.2. To maintain comparability with the 2SLS results in Section 3, we use a result from Blandhol et al. (2022) to calculate the weights on the compliance groups implied by (1)-(2), and report effects using these weights. Second, we study a number of policy-relevant treatment effects, which imply different weights on the effects of each compliance group.

6 Results

6.1 Treatment effects vary over the range of admissible selection models

As discussed in the previous section, the observed information on judge treatment propensities is not sufficient to point-identify the selection equation. As a result, many objects of interest are only partially identified. We illustrate this in Figure 2, which shows the 2SLS-weighted effect of conviction relative to dismissal ($\Delta_{n \to c}$ from (6)) on the number of charges filed over the next five years for different values of ρ. When ρ is small, conviction has almost no effect on charges. As ρ gets larger, however, and the selection model gets closer and closer to a single-index model, the estimated effects rise, approaching 0.07. For $\rho{=}1$, which corresponds to the single-index model, however, the estimated effect shoots down to a statistically significant -0.18. Since the data do not provide any guidance on which ρ is the correct one, we conclude that $\Delta_{n \to c}$ is between -0.18 and 0.07. Incorrectly imposing a relationship between the unobservables presents a real risk; for example, if we had simply used a single-index model of treatment, we would have concluded that convictions substantially decrease future crime even though the data are equally consistent with positive effects.

6.2 Model fit

We approximate the MTRs using the flexible polynomials discussed in Section 5.5. These functions impose that the potential outcomes are smooth in the indices that govern selection, and place implicit restrictions on the slopes of potential outcomes across the space of u.

Our methodology allows us to directly assess whether these MTRs accurately approximate parameters of interest, such as the 2SLS estimates $\widehat{\beta}^{\text{2SLS}}_{incar}$. For each ρ, we denote the estimated MTRs $\widehat{\alpha}$ and the corresponding model-based incarceration 2SLS estimate as $\beta^{\text{2SLS}}_{incar}(\widehat{\alpha})$, and bootstrap the null distribution of $\widehat{\beta}^{\text{2SLS}}_{incar} - \beta^{\text{2SLS}}_{incar}(\widehat{\alpha})$ using i.i.d. draws from the set of cases.

Figure A1 shows the estimated p-values for models of both the number of charges and an indicator for any charge within 5 years over the range of ρ. The 2SLS estimates are similar to their structural analogs; for example, the 2SLS estimate for any charge is -0.058 while the structural estimate $\beta^{\text{2SLS}}_{incar}(\widehat{\alpha})$ with $\rho{=}0$ is -0.044.

Since the choice model is not correct for all values of ρ, we expect that the model might reject for some values. The p-values are close to zero for $\rho{=}1$, consistent with the single-index model being too restrictive. For other values of ρ, the p-values are higher, in the range of 0.20 for the binary outcome and 0.03 for the continuous outcome. We conclude that the polynomials do a good job of approximating the underlying MTRs.

6.3 Decomposing the 2SLS estimands

Except in restrictive choice models, the existence of multiple treatments poses complications for the interpretation of binary 2SLS estimates. As we discuss in Section 3.2, the incarceration 2SLS estimand, for example, is a weighted combination of the effect of incarceration relative to both conviction and dismissal, as well as monotonicity and exclusion violations. While researchers commonly make assumptions that preclude some or all of these effects, there is typically no empirical guide to evaluating their existence and magnitude.

Our approach instead allows us to directly estimate each of these components and evaluate the threat that exclusion and monotonicity violations pose to interpretation. The following equation shows the decomposition of the incarceration 2SLS estimate on the number of charges over the five years following case filing using (6). For each treatment effect and weight, we report the upper and lower bounds on the model estimates across ρ, except for the model-based 2SLS estimand, for which we report the $\rho{=}0$ estimate.

$$\underbrace{-0.241}_{\beta^{2SLS}_{incar}} = \underbrace{[0.977,\ 1.000]}_{\phi_{c\to p}}\underbrace{[-0.246,\ -0.202]}_{\Delta_{c\to p}} + \underbrace{[0.000,\ 0.054]}_{\phi_{n\to p}}\underbrace{[0.220,\ 2.523]}_{\Delta_{n\to p}} + \underbrace{[-0.034,\ 0.000]}_{\text{mono. violations}} + \underbrace{[-0.003,\ 0.005]}_{\text{exclusion violations}} \tag{13}$$

The equation shows that 2SLS does a remarkably accurate job of estimating the causal effect of incarceration. It also reveals which fallback option is relevant: for each value of ρ, the weight is almost entirely on $\Delta_{c\to p}$, the effect of incarceration relative to conviction. The weight on the $n\to p$ effect is always smaller than 0.054, and so the weighted effect of incarceration relative to the alternatives $\beta_{(n,c)\to p}$, at $[-0.213, -0.202]$,[25] almost perfectly coincides with the estimated $c\to p$ effect of $[-0.246, -0.202]$. Furthermore, the exclusion and monotonicity terms are tightly bounded around zero, meaning that the 2SLS estimate almost exactly reflects $\beta_{(n,c)\to p}$.

We also decompose the conviction 2SLS estimand into its constituent parts. This estimate results from instrumenting for conviction (either p or c) with judge indicators. In our choice model there are no $c\to n$ or $p\to n$ defiers, and so the conviction 2SLS estimand can be decomposed into a weighted combination of (1) the treatment effect of p relative to n, (2) the treatment effect of c relative to n, and (3) exclusion violations between c and p:

$$\underbrace{0.189}_{\beta^{2SLS}_{convic}} = \underbrace{[0.000,\ 0.237]}_{\phi_{n\to p}}\underbrace{[0.486,\ 2.897]}_{\Delta_{n\to p}} + \underbrace{[0.768,\ 1.005]}_{\phi_{n\to c}}\underbrace{[-0.189,\ 0.017]}_{\Delta_{n\to c}} + \underbrace{[0.111,\ 0.311]}_{\text{exclusion violations}} \tag{14}$$

This decomposition reveals that the conviction 2SLS estimate of 0.189 does not accurately reflect the causal effect of conviction on future criminal behavior, nor any other causal effect. While the majority of the weight is on $n\to c$ compliers, in contrast to the 2SLS estimate these effects are mostly negative, ranging from -0.189 to 0.017. The effect for the $n\to p$ group is larger, at $[0.486, 2.897]$, but the weight on these compliers is small enough that the combined $\beta_{n\to(c,p)}$ effect is no larger than 0.078.[26] The exclusion violations are bounded between 0.111

[25] This estimand is shown in Equation 11.

[26] To see this, note we can rewrite $\beta^{2SLS}_{convic} = \beta_{n\to(c,p)} + \text{exclusion violations}$.

and 0.311 and statistically significant. Thus, a naive use of 2SLS would overstate the increases in crime resulting from conviction.

Importantly, the forms of bias we find in the 2SLS estimates are not easily corrected by more typical methods. For example, Panel C of Table 2 shows that simultaneously instrumenting for incarceration and treatment using 2SLS barely changes the estimated effect of conviction from 0.084 to 0.088. One would therefore continue to erroneously over-estimate the effect of convictions on future criminal behavior

6.4 Characterizing the compliers

The last section decomposed the effect of incarceration into the weighted sum of two counterfactual-specific effects: incarceration relative to dismissal (d) or conviction (c). These effects differ profoundly, with one reducing recidivism and the other *increasing* it. While this difference might arise from the fallback options themselves, it might also reflect differences in the types of individuals in these groups, such as the extent of their criminal record. In this section we more fully characterize the compliers by calculating averages of pre-existing case and demographic variables for each group, using the same complier weights and identification strategy as our treatment effects.[27] By shedding light on who the compliers are for these separate margins, this analysis is also valuable for understanding which cases judges disagree upon and, hence, which individuals may be marginal to local policies. We investigate differences in fallback options in later sections.

Table 3 reveals that the individuals in different compliance groups face different types of charges and have varying criminal histories. Panel A focuses on the severity of the case, which we measure using the mean sentence that defendants receive when they are charged with the same offense in the same court. A longer expected sentence corresponds to an offense with a longer statutory penalty or a higher risk of conviction, both of which indicate a more serious crime. Among felony cases, incarcerated defendants whose cases would otherwise be dismissed ($n \to p$ compliers) face an expected $[277, 480]$ days behind bars, while incarcerated defendants with a conviction fallback ($c \to p$ compliers) face only $[207, 225]$ days. Similarly, Panel B shows that $n \to p$ compliers also have longer criminal records than $c \to p$ compliers, with $[2.64, 4.73]$ and $[2.55, 2.72]$ prior offenses, respectively.[28]

We observe the same pattern among misdemeanor defendants in column (2), where $n \to p$ defendants are charged with more serious offenses and have a longer criminal record. In particular, their expected sentence is $[11, 14]$ days versus only $[8, 10]$ days for $c \to p$ compliers, and they have at least 2.08 prior offenses while $c \to p$ compliers have $[1.67, 2.11]$. We view this as consistent with our model of court decision-making, where $n \to p$ compliers are more likely

[27] To estimate these complier averages we parameterize the mean characteristics C for each marginal defendant as a polynomial approximation of $E[C|U_1 = u_1, U_2 = u_2, X = x]$. Since this differs from our MTR specification only in being constant across treatments, we can then use our baseline approach to estimate mean characteristics using the same complier weights. Because of the skew in the characteristics we describe, we cap outcomes at the 99th percentile for each type of case and use linear rather than quadratic MTRs, which are more sensitive to outliers.

[28] While the identified sets overlap, for each value of ρ the $n \to p$ group has more prior offenses than the $c \to p$ group.

to have been charged with serious offenses that come with long prison sentences but are only marginally guilty.

We also characterize the $n \to c$ compliers. Compared to $n \to p$ defendants, who are on the margin between dismissal and incarceration, they have fewer previous charges ($[1.57, 2.18]$ versus $[2.64, 4.73]$ for felony defendants) and a shorter expected sentence ($[175, 242]$ versus $[277, 480]$ days). These differences, which we also see among misdemeanor defendants, are again consistent with our model of judge decision-making: defendants with a marginal case against them are caught between dismissal and incarceration if they have been accused of a more serious crime, and caught between dismissal and a non-carceral sentence when they have been charged with a more minor offense.

Finally, Panel C displays the effect of incarceration on sentence length for each fallback option. Consistent with the more serious offenses that $n \to p$ compliers have been charged with, the marginal incarceration increases sentences by $[390, 1012]$ days for felony defendants whose cases would otherwise be dismissed, but by a more modest $[384, 388]$ for defendants who would otherwise be convicted. In contrast, we see no such difference in effects on sentence length for misdemeanor defendants, and the identified sets overlap. The marginal sentence is also much shorter for misdemeanor defendants, at no more than 44 days, previewing the relatively small role that incarceration will have on defendants relative to conviction for this group.

6.5 Effects of conviction and incarceration on future criminal behavior

Table 4 presents the effects of incarceration and conviction on the total number of future charges (Panel A) and convictions (Panel B) over the five years following case filing. The first column, which like (13) studies misdemeanor and felony defendants together, reveals that nearly all of the variation in the judge instruments shifts defendants between c and p; the weight on the $c \to p$ effect is at least 0.977 across values of ρ. This allows a precise estimate of between 0.202 and 0.246 future charges and $[0.233, 0.278]$ convictions averted for each marginal incarceration.

In contrast, the treatments that lead to a conviction, $\Delta_{n \to p}$ and $\Delta_{n \to c}$, are relatively imprecisely estimated, and we can't reject that there is no effect of either treatment. While this might suggest that convictions are not important determinants of criminal justice outcomes, this is driven by two different factors: relatively low precision for felony cases caused by limited across-judge variation in conviction propensity, and smaller effects of a conviction for defendants who already have a criminal record. When we focus on the populations more likely to be affected by a conviction, and for whom we have more statistical power, we see more precisely estimated and deleterious effects of conviction.

Effects for felony defendants

Column (2) of Table 4 reports the effects of conviction and incarceration for felony defendants. At least 99.1% of the weight is on the $c \to p$ effect, and for this group the table reveals that incarceration reduces the number of future charges by between 0.345 and 0.379 over the five years following filing.

To understand the possibly mechanical role of incapacitation, Figure 3 plots the effects of incarceration on both days spent locked up (whether resulting from the focal or non-focal case) in each year and on the cumulative number of charges filed against the defendant since case filing. Incarceration results in an additional 200 days spent incapacitated in the first year, but this effect quickly fades to only 25 extra days in the third year. The effects on cumulative charges mirror this trajectory, with a big decline in the first year and a smaller decline in the second year. Over the next five years, after the incapacitation effect has faded, there is essentially no further effect of initial assignment to incarceration on the number of charges. This indicates that incapacitation rather than changes to post-release behavior likely explains nearly all of the effects of incarceration relative to conviction, consistent with similar effects documented in prior work (Rose and Shem-Tov, 2021).[29]

Unfortunately, the effects of conviction on future behavior for felony defendants are relatively imprecisely estimated because of limited variation in conviction propensities, and we cannot reject a null of no effect on future charges (Panel A) or convictions (Panel B). This judicial conformity suggests that for felony courts, reducing conviction rates is likely to require interventions before the judges are assigned the cases, possibly at the level of the prosecutor making charging decisions or changes to law (Agan, Doleac and Harvey, 2022).

Effects for misdemeanor defendants

As discussed in Section 6.4, misdemeanors carry much shorter sentences than felonies. This results in relatively little incapacitation for incarcerated compliers who would otherwise be convicted. Figure 3 shows that even in the first year after case filing, incarceration for misdemeanor $c \rightarrow p$ compliers results in only 30 additional days behind bars. In the second year, the effect is indistinguishable from zero. Consistent with incapacitation being the driving factor for $c \rightarrow p$ defendants—all of whom are convicted no matter their judge assignment—we observe no effect of incarceration on the number of future charges for any of the years following the case.

We can rule out impacts of misdemeanor incarceration larger than 0.15 crimes in either direction, or about 10% relative to the dependent variable mean number of crimes, at the 95% level.[30] We can thus clearly reject that the impacts are the same for felony incarceration and reject medium-sized increases in future charges. Given that millions of people are jailed each year (Zeng, 2020), this causal estimate provides new and informative results relevant to the impacts of the criminal justice system more broadly.

Despite the limited effects for misdemeanor incarceration, convictions could still affect individuals' future outcomes. A criminal record might make it harder to gain employment (Pager, 2003), and since police officers and prosecutors can see the record of convictions, future criminal justice system involvement might be more likely to result in charges. This might be especially true for $n \rightarrow p$ compliers, who have been accused of more serious crimes

[29] Table 3 reports that incarceration results in an average sentence of $[384, 388]$ days for $c \rightarrow p$ compliers, implying that a year's sentence averts $[0.345, 0.379]/([384, 388]/365) = [0.325, 0.360]$ new offenses.

[30] Column (3) in Table 2 shows the mean number of charges over the 5-year period after case filing for misdemeanor defendants.

but whose guilt is marginal.

Panel A of Figure 4 plots the effects of incarceration relative to dismissal for the $n \to p$ compliers. The sentence length is relatively short; while days incapacitated increases by about 60 in the first year, there is no effect in any subsequent year. However, despite the modest degree of incapacitation, the number of future charges *increases* as a result of treatment. By year 5, being assigned to p rather than n has increased the number of future charges by $[0.563, 4.101]$, and the number of future convictions by $[0.786, 4.987]$.

To investigate the relative roles of incapacitation versus receiving a criminal record, we use our model to decompose the effect on future charges for the $n \to p$ compliers into the effect of moving from n to c, and the effect of moving from c to p. Using superscripts to denote the compliance group and subscripts for the effect, this reveals that

$$\underbrace{\underset{(0.094,6.292)}{[0.563, 4.101]}}_{\Delta_{n\to p}} = \underbrace{\underset{(0.80,6.24)}{[1.88, 3.93]}}_{\Delta^{n\to p}_{n\to c}} + \underbrace{\underset{(-3.04,3.40)}{[-1.33, 0.191]}}_{\Delta^{n\to p}_{c\to p}}$$

In other words, the $n \to p$ effect is entirely driven by conviction relative to dismissal, which increases the number of future charges by somewhere between 1.88 and 3.93. There is no effect of jail time conditional on conviction.

We also study the effect of conviction for $n \to c$ compliers. Since this group has typically been accused of somewhat less serious crimes and would likely be less impacted by social stigma than $n \to p$ compliers, there may be less scope for a conviction to affect future behavior. This is exactly what we find: Panel B of Figure 4 reveals no statistically significant effect of conviction for $n \to c$ compliers on cumulative charges in any of the first seven years following case filing. However, it *does* increase the number of future convictions by $[0.222, 0.557]$ after five years (Panel B of Table 4). This is much smaller than the $n \to p$ effect of $[0.786, 4.987]$, consistent with the less serious offenses that $n \to c$ compliers face.

The positive effects of conviction on future crime have several interesting implications. First, the high rates of misdemeanor conviction in the US imply that the increases in crime and future conviction we find may be particularly important. Second, the larger effects on future convictions than on future charges are consistent with police, prosecutors, and judges treating defendants more harshly in *future* cases as a result of a past conviction, and suggest that initial inequities in criminal justice contact can lead to persistently differential treatment. Finally, our results suggest that convictions—rather than incarceration or some other aspect of criminal prosecution—may be the key mechanism behind recent work showing that a prosecutor's decision to proceed with a case increases future crime (Agan, Doleac and Harvey, 2022).

More evidence on the central role of convictions in misdemeanor cases comes in column (5) of Table 4. In this column we restrict to misdemeanor defendants who have never been convicted of a felony offense, and so might be more profoundly affected by a conviction.[31] Consistent with this, we see that the lower bounds on the $n \to p$ and $n \to c$ effects are larger

[31] We also examine the effect of criminal justice sanctions for never-previously-convicted *felony* defendants in column (4). However, given the imprecision of our estimated conviction effects for felony defendants, we cannot rule out even relatively large changes in behavior.

for this group than for all misdemeanor defendants. Weighting by the relative size of the two groups, conviction causes an additional $[0.165, 0.817]$ charges to be filed over the next five years. This is composed of $[0.851, 3.336]$ for $n \to p$ defendants, and $[0.165, 0.351]$ for the $n \to c$ compliers, with the smaller effect for the noncarceral sentence reflecting the less serious nature of the charges. The effects on the number of future convictions (Panel B) are slightly larger than those on future charges for both compliance groups, again consistent with more punitive behavior from future criminal justice officials.[32]

6.6 Discussion

The results in this paper help clarify the respective effects of conviction and incarceration, and unify the existing literature. As we discuss in the introduction, recent research has found very different effects of these treatments, with incarceration relative to conviction on felony charges reducing future crime and the conviction-related impacts of criminal prosecution increasing it. Thus, although both of these treatments increase the intensity of criminal justice contact and the associated degree of social stigma, they have opposite effects on recidivism. Whether this difference in effects is broadly generalizable across the United States, or stems from methodological or external validity distinctions, is not clear from existing work.

Our study provides a simple explanation for this pattern of results. Using a single research design and setting, we replicate the qualitative patterns of the prior literature. Conviction on minor offenses—like those studied in Agan, Doleac and Harvey (2022) and Mueller-Smith and Schnepel (2021)—increases future crime, while incarceration on felony charges decreases future crime. Since felony incarceration affects recidivism only during the period of incapacitation, and there is no effect of relatively short misdemeanor sentences on future crime, we view these results as most consistent with incarceration affecting future crime only through incapacitation. The deleterious long-term effects of criminal prosecution therefore appear to arise mostly from the conviction and the resulting criminal record. Consequently, heterogeneity across locations, research designs, or other contextual features do not seem to contribute to the disparate results in the field.

Our novel estimates of the effect of misdemeanor incarceration on recidivism also shed light on a recent literature that has studied the effect of misdemeanor pre-trial detention (Gupta, Hansman and Frenchman, 2016; Dobbie, Goldin and Yang, 2018; Heaton, Mayson and Stevenson, 2017). This literature has typically found that pre-trial detention increases recidivism, but it is unclear if this is because the detention itself is criminogenic or because it indirectly affects recidivism by increasing conviction rates. Our finding that misdemeanor convictions increase future recidivism—but that incarceration does not—neatly explains these results.

Finally, this work complements two previous papers that use similar data to study the externalities of incarceration on children and siblings (Norris, Pecenco and Weaver, 2021) and

[32] Charges that do not lead to a conviction will typically be visible to police in the same local area. However, convictions will be visible more widely and are regarded as more serious than simply being charged (Agan, Doleac and Harvey, 2022).

the impacts of incarceration on earnings for defendants (Garin et al., 2023). We find that in a 2SLS regression with incarceration as a treatment, the monotonicity and exclusion violations are small, strengthening the causal interpretation of this prior work. Our approach also provides further evidence that the 2SLS estimands reflect the effect of incarceration relative to a counterfactual of conviction (instead of dismissal).

6.7 Policy effects

In the previous section, we reported effects of conviction and incarceration for individuals weighted by their response to the changes in the judge instruments. While these weights provide a useful benchmark, they do not necessarily capture populations affected by particular policy changes (Heckman and Vytlacil, 2005). In this section, we directly study the effects of policies that change judge behavior, including policies that induce responses along both conviction and incarceration margins. As our previous 2SLS-weighted results concord closely with the qualitative patterns of the literature, the policy impacts we study may also be generalizable.

We focus on two types of policy reforms. First, *local* changes make small reductions to judge thresholds $g_1(z)$ and $g_2(z)$. We view these as approximating what would happen if judges became more lenient with respect to conviction or sentencing, respectively. Greater conviction leniency could arise from a higher evidentiary standard or greater willingness of prosecutors to drop cases where they viewed the evidence as marginal. Sentencing leniency could arise from reforms to sentencing guidelines that made probation the presumptive sentence for a wider range of defendants. We estimate the effects of greater conviction leniency by adjusting each judge's g_1 threshold by 0.01, reassigning treatment with the new threshold, and using the estimated MTRs to predict outcomes under the counterfactual policy. Similarly, we predict the effect of greater incarceration leniency by decreasing g_2 by 0.01 for each judge and estimating outcomes under the new treatment assignment.

Second, we study *global* policy changes that eliminate either conviction or incarceration. We estimate the effect of these policies by reducing the g_1 (respectively, g_2) threshold to zero for each judge and calculating outcomes under the new treatment assignment with the estimated MTRs. Importantly, these policy effects do not take into account general equilibrium responses and so may overstate the benefits of these policies. Nonetheless, they help illustrate the possible effects of larger, non-marginal policy changes.

Effects for felony defendants

Panel A of Table 5 reports the effects of each policy change for felony defendants. Consistent with the 2SLS effects, both marginal increases in sentencing leniency as well as banning incarceration increases the number of offenses committed by the defendants over the following five years. The first row reports that the marginal defendant who would be spared incarceration by increasing sentencing leniency will commit an additional 0.356 to 0.66 crimes as a result of the policy change, and be convicted in an additional 0.256 to 0.609 cases. Similarly, banning incarceration would reduce the incarceration rate from 28.9% to 0% and increase the number of future charges by $[0.085, 0.295]$ per defendant (or $[0.294, 1.021]$ per affected person) even before accounting for any general equilibrium effects.

Although these results mean that expanding sentencing leniency would increase crime, they do not consist of a full cost-benefit analysis. In particular, prisons are expensive; a year of incarceration in Ohio costs approximately $26,500 in 2015 dollars (Mai and Subramanian, 2017). We estimate that the marginal incarceration in this policy change is for $[500, 513]$ days, implying a cost per averted future charge of somewhere between $55,002 and $104,621.[33] This is higher than the $20,472 estimated cost of the average crime, although this comparison abstracts from the relationship between averted future charges and averted future crime (Miller et al., 2021).[34]

Effects for misdemeanor defendants

Panels B and C of Table 5 study the effect of increased leniency for misdemeanor defendants and never-previously-convicted misdemeanor defendants, respectively. In line with the small incapacitation effects we observed in Section 6.5, the bounds on the effect of increased sentencing leniency staddle zero for both populations. While not directly-crime reducing, this type of leniency would still reduce jail populations without large increases in future crime.

Expanded leniency in conviction is more promising. The second rows of Panel B and C report the effect of marginally increasing conviction leniency for misdemeanor defendants. They reveal that on average it would have no effect on the number of future charges for the overall misdemeanor defendant (bounds of -0.112 to 0.215), and slightly (although not statistically significantly) reduce charges by 0.022 to 0.247 for the never-previously-convicted population. The benefits are slightly larger in terms of avoiding future convictions; the marginal never-previously-convicted beneficiary would see 0.224 to 0.489 fewer convictions due to this policy change.

These benefits are somewhat smaller than would be expected from the 2SLS estimates. A closer examination of the compliance patterns reveals why. For each population, 80-100% of the beneficiaries of increased conviction leniency would otherwise be convicted but not incarcerated. As we discussed in Section 6.5, the benefits of case dismissal are smaller for the $n \to c$ population than the $n \to p$ population. This substantially reduces the benefits of expanding leniency across the board, and suggests that more effective policy reforms would focus on defendants accused of more serious offenses who would otherwise be incarcerated, since the benefits disproportionately accrue to this group.

We also study the effect of larger increases in conviction leniency. While these estimates do not account for changes in general deterrence, they suggest that the effect of case dismissal for non-marginal defendants are substantially higher. For example, dismissing all cases would reduce the number of future convictions by $[0.858, 0.869]/0.53 = [1.62, 1.64]$ for the average affected misdemeanor defendant, compared to $[0.052, 0.449]$ for the defendant affected by a marginal increase in leniency. Policies that encourage diversion for misdemeanor defendants,

[33] $26500 \times ([500,\ 513]/365)/[0.356,\ 0.660] = [55002, 104621]$.

[34] Miller et al. (2021) reports costs in 2017 dollars; we use the CPI to adjust to 2015 for comparability.

particularly those without a long criminal record, therefore can potentially decrease crime and future involvement with the criminal justice system.

Effects of homogenizing judge behavior

Finally, we study the effect of homogenizing judge behavior. A large literature has noted disparities in conviction and incarceration rates across judges, and Ohio has moved to make more sentencing data available to judges with the explicit goal of reducing these disparities (Ohio Criminal Sentencing Commission, 2021). Harmonizing judge sentencing behavior can have important impacts on outcomes if their effects are asymmetric—for example, if harsh punishments by the least lenient judges cause some people to get caught up in recurring interactions with the criminal justice system.

We implement this counterfactual by assigning all judges in each x cell the same g thresholds, picking them so that the overall treatment shares remain the same. We then study the effects on outcomes relative to the status quo. Table 5 reveals that these policies would have only a limited effect on outcomes. Across courts and types of defendants, there is no effect of homogenizing judge behavior on any of the outcomes, although for misdemeanor defendants the bounds typically barely include zero. This suggests policies to harmonize judge behavior will not directly affect recidivism, although there are likely benefits to increasing certainty in the legal system.

6.8 Robustness

Measurement horizon

Our main results report effects after 5 years. We choose this horizon because it is long enough that several years have elapsed since most defendants have been released (see Figure 3), but short enough that the sample size remains large. As a robustness exercise, we also estimate these effects after 7 years. Table A4 contains the incarceration 2SLS-weighted effects (analogous to Table 4), and Table A5 reports policy effects after 7 years (analogous to Table 5). While the edges of the bounds are sometimes slightly smaller or larger, the substantive conclusions are unchanged.

Binary recidivism

Our baseline approach has been to measure the effects of treatment on the number of future charges and convictions. However, some recent work has also studied the effect on binary recidivism (Jordan, Karger and Neal, 2023). For comparability, in Table A1 we re-estimate our main results (Table 4) using binary measures of recidivism. The results are qualitatively similar: incarceration tends to decrease future criminal justice involvement, while conviction tends to increase it, particularly for first-time misdemeanor defendants.

Measure of prior criminal justice contact

One key population of interest in this study is defendants with no prior felony conviction, who we find are more strongly affected by convictions (and the resulting criminal record) than the average defendant. In Table A2 we report results using the smaller sample of defendants with no prior felony or misdemeanor convictions. The results are slightly less precise but largely comparable. Summarizing the conviction effect $\Delta^*_{n\to c}$ as the effect of c relative to n for the weighted $n\to c$ and $n\to p$ complier groups, the effect on future charges for never-felony-convicted misdemeanor defendants is $[0.165, 0.813]$ compared with $[0.173, 0.712]$ for defendants who had never been convicted of either a felony or misdemeanor.

Weights used in estimand

Our main results (Table 4) use the implied weights from a regression instrumenting for incarceration in constructing the estimand. In Table A3, we report the same effects using the implied weights from a regression instrumenting instead for conviction. This aggregation allows us to study a population more responsive to differences across judges with varying conviction propensities, and provides a sense of heterogeneity in results when compared with the results using incarceration weights. Comparing effects on charges within 5 years, of most note, the effect of felony incarceration relative to conviction is now a positive though insignificant [0.069, 0.133], and we can reject that this estimate overlaps with the incarceration 2SLS-weighted estimate of [-0.379, -0.345]. This finding highlights important heterogeneity in the effects of incarceration and is inconsistent with constant treatment effects in this population. While we also see differential responses in effects of p relative to n, the effects of c relative to n appear consistent across the complier populations.

Parametric family for distribution of unobservables

We also consider alternative distributions for the unobservables. Our baseline model uses a copula based on the bivariate normal distribution. However, we also consider using a copula that is instead based on a distribution with logistic marginals and that allows for correlation between the two dimensions (Ali, Mikhail and Haq, 1978). This is a useful comparator because of the prevalence of the logistic distribution in standard models of discrete choice. As in our baseline analysis, we allow the correlation parameter to take values between 0 and 1.

Table A6 shows the effects of conviction and incarceration on the number of future charges and convictions, analogously to Table 4. The results are similar, although the bounds are somewhat tighter using the AMH copula. For example, conviction relative to no punishment increases the number of future charges by $[0.165, 0.351]$ for never-convicted misdemeanor defendants in our baseline analysis, but by $[0.293, 0.343]$ using the AMH copula.

7 Conclusion

Estimating the causal effect of criminal justice sanctions is difficult due to non-random assignment of treatments. Examiner designs use variation from randomly assigned judges as instrumental variables to study the effect of a particular sanction, such as incarceration, on individual outcomes. However, simple 2SLS models typically cannot account for judges choosing between three or more treatments—such as dismissal, conviction and incarceration—thereby biasing estimates.

We build a new framework to handle these and other similar situations that feature discrete instruments and multiple treatments. We introduce a choice model appropriate for judge settings, link it to an equivalent selection model, and develop a novel estimation framework to recover 2SLS-weighted combinations of treatment effects stripped of monotonicity and exclusion violations. We then go beyond these estimands to isolate the component treatment effects and to extrapolate to well-defined alternative policies. Our approach, which requires only examiner instruments for identification, allows for more flexible substitution patterns—and thus, more credible estimates—than existing alternatives.

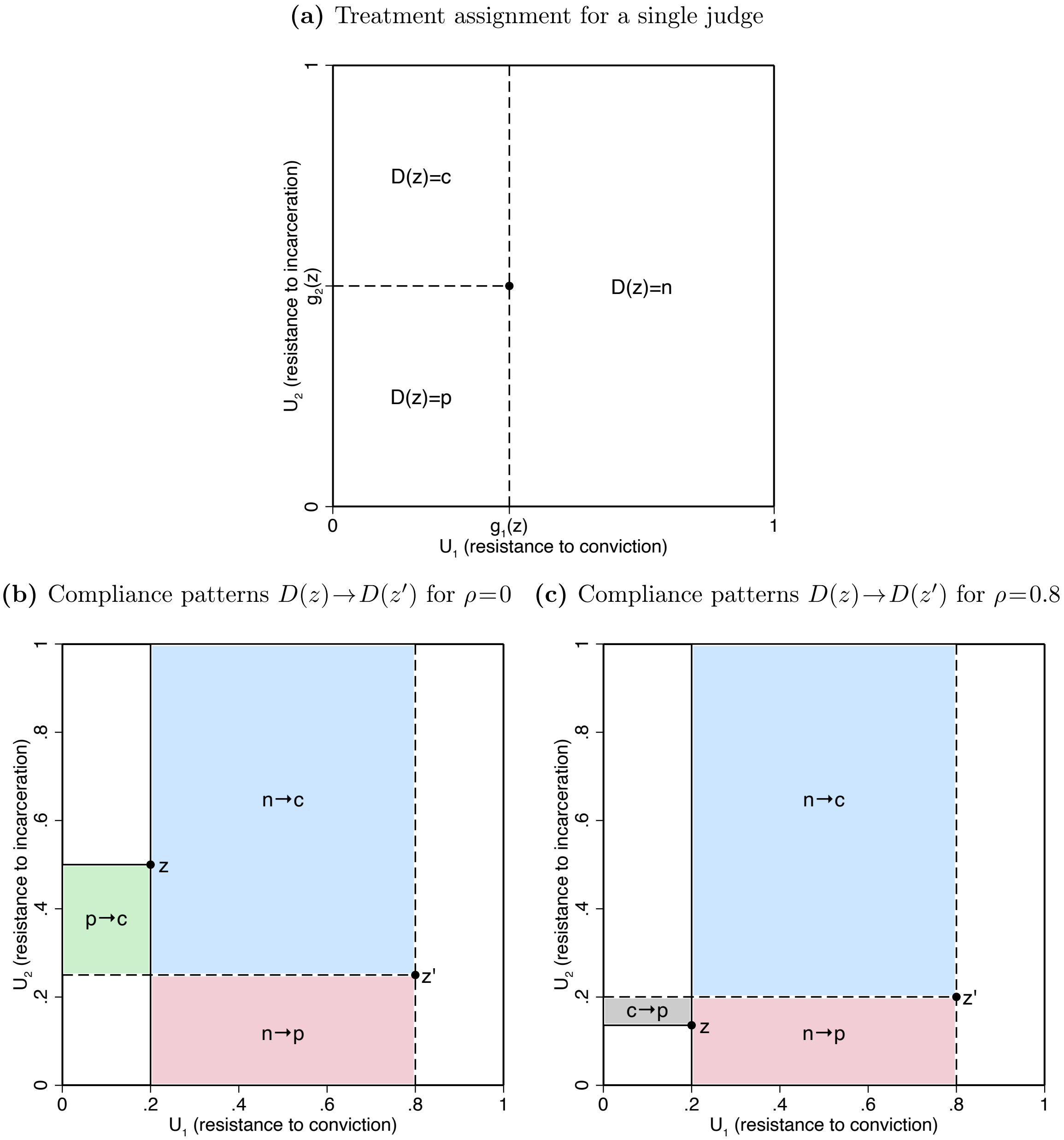
We use this model to study the effect of conviction and incarceration in the three largest counties in Ohio. We reconcile a string of recent results in the economics of crime literature studying these treatments. Consistent with prior work, we find that incarceration decreases future crime through an incapacitation effect, while misdemeanor convictions *increase* subsequent criminal justice involvement. One contribution of this paper is to show that these results hold in a single setting and research design, assuaging external validity concerns of prior work. We go beyond this by decomposing previously studied treatment effects into margin-specific effects and by providing new estimates of less-studied sanctions of the criminal justice system, including misdemeanor incarceration. In so doing, we highlight the important differences in the effect of sanctions in misdemeanor and felony courts, emphasizing the importance of further work on this topic.

Most substantively, this paper has important implications for policy. We find that courts could implement reforms that both increase leniency and decrease crime, particularly if they target misdemeanor defendants with short criminal records and who face relatively serious charges.

The approach used in this paper may be useful in other settings. For example, foster care caseworkers first decide whether to remove a child from their parents, and then whether to place them with relatives or non-relatives. Other settings may require slightly different choice models. Disability examiners may affect claimants through both time to decision and benefits receipt (Autor et al., 2011), which could be modeled as two separate dimensions that examiner's behave monotonically over. Corporate bankruptcy judges decide between Chapter 11, Chapter 7, and case dismissal, which may be best described by a multinomial decision model. Our approach of partially identifying the first stage and estimating outcomes using marginal treatment effects can still be applied in these cases with an appropriate adjustment to the choice model.

Figures

Figure 1: Illustration of threshold crossing model for judge decisions



(a) Treatment assignment for a single judge

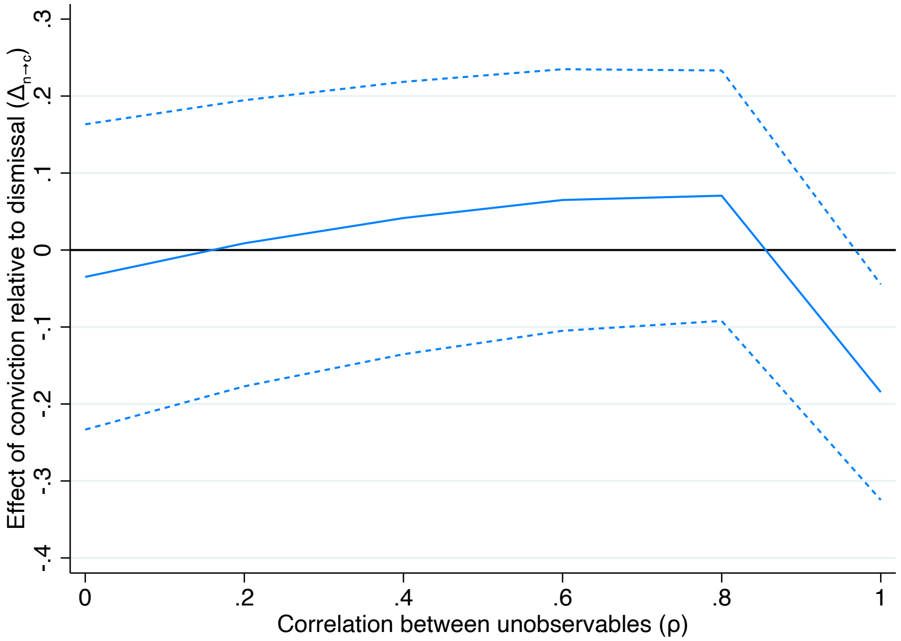
(b) Compliance patterns $D(z)\to D(z')$ for $\rho=0$

(c) Compliance patterns $D(z)\to D(z')$ for $\rho=0.8$

D(z)=c
D(z)=p
D(z)=n
g₂(z)
g₁(z)
U₂ (resistance to incarceration)
U₁ (resistance to conviction)
0
1
p→c
n→c
n→p
c→p
z
z'
0
.2
.4
.6
.8
1

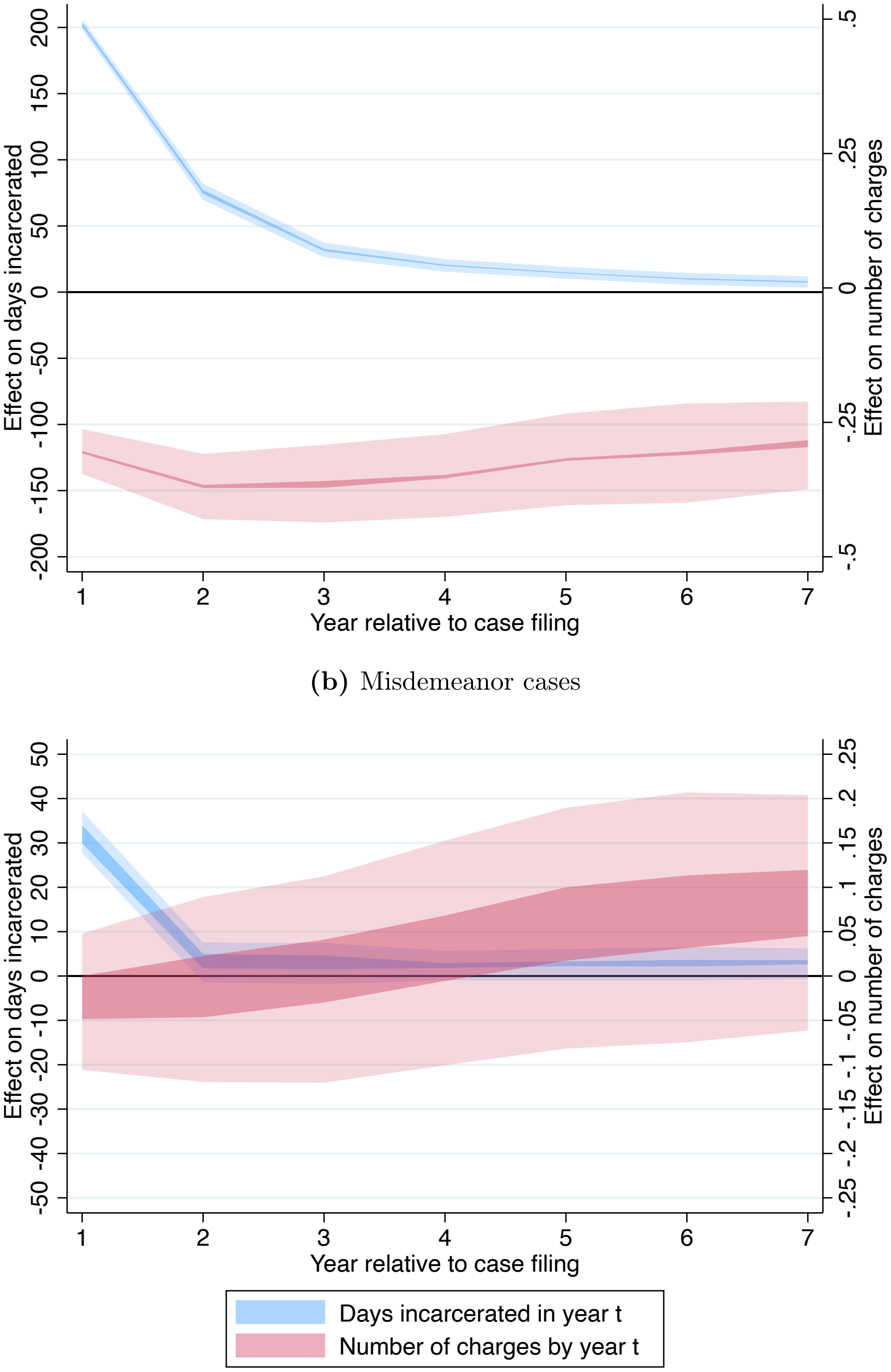
Panel A illustrates treatment assignment decisions for a single judge. Panels B and C show the compliance patterns $\big(D(z)\to D(z')\big)$ for judges z and z' for $\rho \in [0, 0.8]$. Judge z assigns 80, 10, and 10% of defendants to treatment n, c, and p, respectively. Judge z' assigns 20, 60, and 20%. As ρ changes so does the share of each response group; in particular there are $p\to c$ compliers for $\rho = 0$ but not $\rho = 0.8$. Similarly there are $c\to p$ compliers for $\rho = 0.8$ but not $\rho = 0$. The white regions in Panels B and C denote values of U where changing the judge from z to z' does not affect the realized treatment.

Figure 2: Effect of conviction relative to dismissal ($\Delta_{n\to c}$) on 5-year number of charges, by value of ρ



This figure shows the estimated effect of conviction relative to dismissal on the number of subsequent charges over the five years following the focal case filing. We display these estimates for $\rho \in [0, 0.2, 0.4, 0.6, 0.8, 1]$.

Figure 3: Effect of incarceration relative to conviction on days incarcerated and number of charge

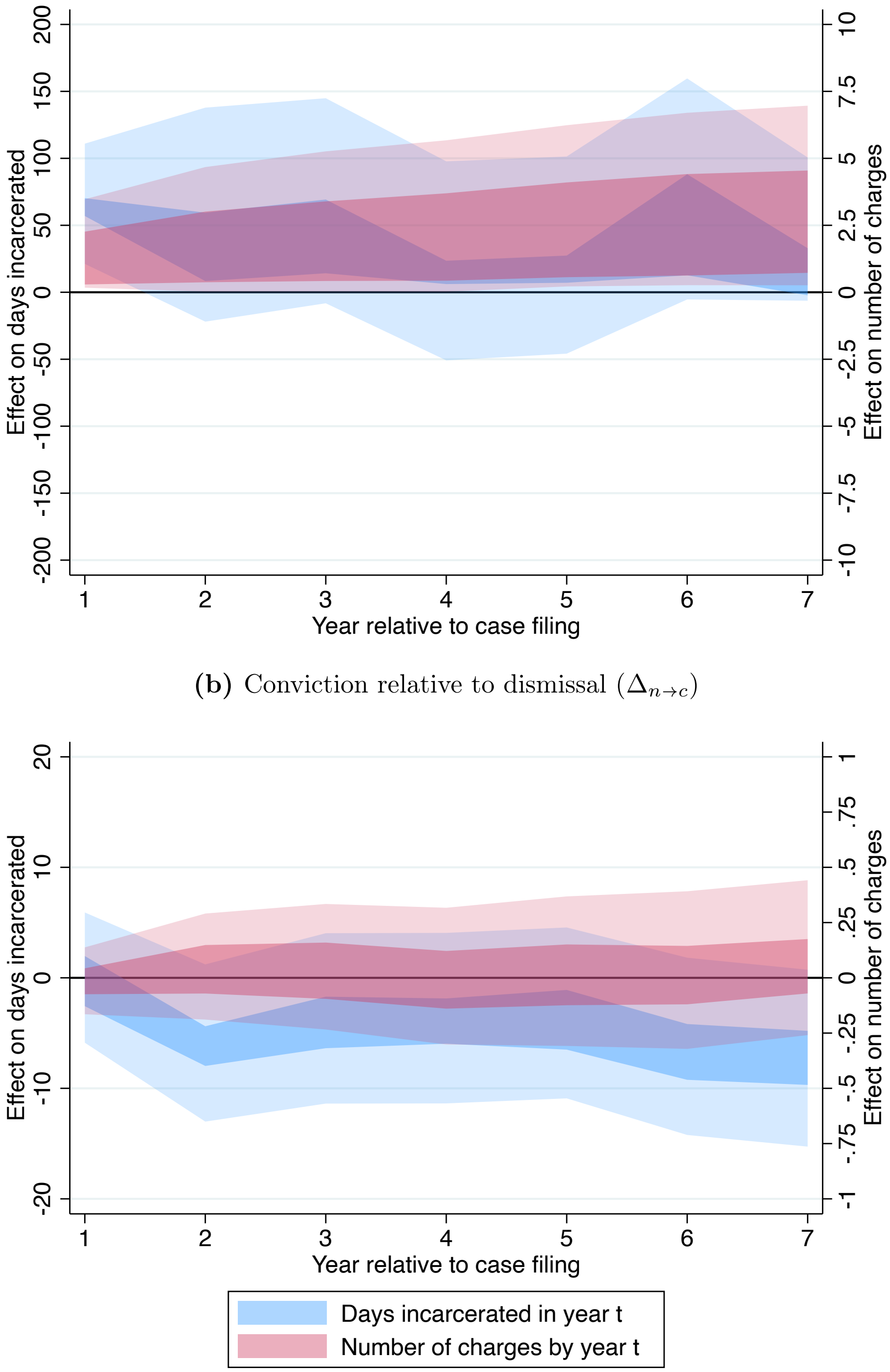


(a) Felony cases

(b) Misdemeanor cases

This figure shows the estimated effects of incarceration relative to conviction ($\Delta_{c \to p}$) on the number of days spent incarcerated and number of new charges in each year after case filing. The darker areas denote the range of estimates arising from choice models with selection parameters $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The lighter areas denote the edge of 95% confidence intervals for the endpoints estimated with Bei (2023).

Figure 4: Effects of conviction on days incarcerated and number of charge for misdemeanor defendants



(a) Incarceration relative to dismissal ($\Delta_{n\to p}$)

(b) Conviction relative to dismissal ($\Delta_{n\to c}$)

This figure shows the estimated effects of incarceration relative to dismissal ($\Delta_{c\to p}$) and conviction relative to dismissal ($\Delta_{n\to c}$) on the number of days spent incarcerated and number of new charges in each year after case filing. Restricted to misdemeanor defendants only. The darker areas denote the range of estimates arising from choice models with selection parameters $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The lighter areas denote the edge of 95% confidence intervals for the endpoints estimated with Bei (2023).

Tables

Table 1: Defendant characteristics and judge severity

	Mean char. by court			Rel. with judge propensities	
	All	Felony	Misd.	Incar.	Convic.
		Panel A: Defendant characteristics			
Male	.77	.80	.75	.0084	.037*
				(.008)	(.022)
White	.39	.34	.43	-.00077	-.023
				(.010)	(.026)
Age	31.99	32.24	31.73	-.14	-.71
				(.224)	(.566)
Drug crime	.29	.34	.23	-.016	.031
				(.010)	(.023)
Violent crime	.19	.15	.22	.0074	-.0036
				(.008)	(.021)
Property crime	.29	.38	.19	.013	-.036
				(.010)	(.023)
Sex crime	.05	.04	.06	.0033	.0011
				(.005)	(.012)
Family crime	.14	.07	.21	-.00097	.028
				(.006)	(.020)
Other crime	.28	.27	.28	-.0097	.013
				(.010)	(.024)
Offense mean sentence (days)	99.66	193.02	6.97	3	5.7
				(3.688)	(6.314)
Log mean sentence	3.00	4.84	1.16	.014	-.066
				(.025)	(.076)
Number of previous charges	2.17	2.53	1.80	-.088	.29
				(.079)	(.202)
Joint p-value				.76	.35
		Panel B: Treatment outcomes			
Not guilty (D=0)	.30	.13	.47		
Conviction (D=1)	.51	.59	.43		
Incarceration (D=2)	.20	.29	.10		
Sentence cond. on incar. (days)	464.56	614.76	37.71		
Observations	638,684	323,046	315,638		
Defendants	375,255	188,681	186,574		

Columns (1)-(3) report the sample means for all courts, felony courts, and misdemeanour courts corresponding to this characteristic, respectively. Columns (4)-(5) report the coefficient from a regression of the characteristic on judge mean incarceration and conviction severity, respectively. Joint p-value comes from an F-test of joint significance of the characteristics on the instrument. Controls include year by court fixed effects. Cases may include multiple charges of different types, so the sum of types of charges sums to more than 1. Charge sentence measures offense severity by calculating the leave-out average sentence for the most serious charge. Standard errors clustered at the defendant level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: 2SLS effects of conviction and incarceration on 5-year number of charges

				Never prev. convicted	
	All	Felony	Misdemeanor	Felony	Misdemeanor
	(1)	(2)	(3)	(4)	(5)
Panel A: Effect of incarceration					
Incarceration (D_p)	-0.362***	-0.484***	-0.131*	-0.391***	-0.033
	(0.043)	(0.052)	(0.075)	(0.076)	(0.089)
Dependent mean	1.616	1.581	1.653	0.977	0.933
J	455.641	260.824	176.636	142.374	104.340
p-value	0.000	0.000	0.000	0.637	0.041
Observations	638,684	323,046	315,638	143,657	167,258
Panel B: Effect of conviction					
Conviction (D_{cp})	0.084	0.519**	-0.076	0.352	0.040
	(0.108)	(0.203)	(0.128)	(0.215)	(0.127)
Dependent mean	1.616	1.581	1.653	0.977	0.933
J	542.857	352.219	179.481	168.304	104.712
p-value	0.000	0.000	0.000	0.133	0.039
Observations	638,684	323,046	315,638	143,657	167,258
Panel C: Effect of both					
Incarceration (D_p)	-0.363***	-0.475***	-0.128*	-0.382***	-0.031
	(0.043)	(0.053)	(0.075)	(0.078)	(0.090)
Conviction (D_{cp})	0.088	0.315	-0.054	0.151	0.036
	(0.109)	(0.208)	(0.129)	(0.221)	(0.127)
Dependent mean	1.616	1.581	1.653	0.977	0.933
J	455.876	259.098	175.981	142.540	104.504
p-value	0.000	0.000	0.000	0.611	0.034
Observations	638,684	323,046	315,638	143,657	167,258

This table reports IV estimates of the effect of incarceration, conviction, and both on cumulative charges up to 5 years post filing. Columns are split by sample, with column (1) including all cases, column (2) including felony cases, column (3) including misdemeanor cases, column (4) including felony cases for defendants with no prior felony convictions, and column (5) including misdemeanor cases for defendants with no prior felony convictions. The endogenous variables are instrumented with the judge identity and all specifications include court-year fixed effects. Standard errors in parentheses and clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Characterizing the complier groups

			Never prev. convicted	
	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)
Panel A: Complier offense mean sentence (in days)				
Incarceration rel. to conviction $(c \to p)$	[207.2, 224.9]	[8.1, 9.6]	[221.9, 253.5]	[7.6, 9.1]
Incarceration rel. to not guilty $(n \to p)$	[277.0, 479.5]	[10.9, 14.3]	[296.8, 545.6]	[10.6, 13.8]
Conviction rel. to not guilty $(n \to c)$	[175.0, 242.3]	[7.6, 8.9]	[168.2, 242.3]	[7.7, 8.8]
Panel B: Complier mean number of previous charges				
Incarceration rel. to conviction $(c \to p)$	[2.545, 2.717]	[1.670, 2.106]	[0.607, 0.626]	[0.676, 0.897]
Incarceration rel. to not guilty $(n \to p)$	[2.643, 4.726]	[2.075, 2.781]	[0.823, 1.301]	[0.684, 0.862]
Conviction rel. to not guilty $(n \to c)$	[1.572, 2.176]	[1.301, 1.515]	[0.774, 0.860]	[0.500, 0.509]
Panel C: Effect on sentence length (in days)				
Incarceration rel. to conviction $(\Delta_{c \to p})$	[384.0, 388.0]	[20.7, 24.0]	[398.9, 401.9]	[20.5, 23.4]
Incarceration rel. to not guilty $(\Delta_{n \to p})$	[390.3, 1011.8]	[-36.9, 43.7]	[392.1, 482.4]	[-50.3, 42.7]

This table reports complier characteristics for different complier groups, aggregated using the weights from a 2SLS regression with incarceration as the treatment and judge dummies as the instruments. MTRs are approximated by a linear MTRs in u_1 and u_2.

Table 4: Effects of conviction and incarceration on future criminal justice outcomes

				Never prev. convicted	
	All	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)	(5)
Panel A: Number of charges over next 5 years					
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.246, -0.202]	[-0.379, -0.345]	[-0.021, 0.071]	[-0.333, -0.265]	[-0.037, 0.024]
	(-0.322, -0.143)	(-0.461, -0.269)	(-0.147, 0.159)	(-0.426, -0.195)	(-0.122, 0.097)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[0.220, 2.523]	[-1.016, 0.018]	[0.563, 4.101]	[-4.237, -0.068]	[0.851, 3.336]
	(-0.124, 5.192)	(-6.481, 4.449)	(0.094, 6.292)	(-11.125, 2.562)	(0.377, 5.959)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[-0.185, 0.071]	[-0.304, -0.087]	[-0.124, 0.151]	[-0.187, 0.015]	[0.165, 0.351]
	(-0.315, 0.217)	(-0.617, 0.167)	(-0.305, 0.365)	(-0.417, 0.227)	(-0.016, 0.492)
Panel B: Number of convictions over next 5 years					
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.278, -0.233]	[-0.364, -0.323]	[-0.150, -0.063]	[-0.272, -0.190]	[-0.074, -0.013]
	(-0.356, -0.164)	(-0.454, -0.235)	(-0.318, 0.096)	(-0.383, -0.088)	(-0.259, 0.164)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[0.331, 3.054]	[-1.280, -0.063]	[0.786, 4.987]	[-0.949, 0.278]	[1.166, 3.870]
	(0.026, 6.159)	(-7.947, 5.386)	(0.073, 8.076)	(-9.108, 7.211)	(0.302, 7.333)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[0.105, 0.358]	[-0.379, -0.033]	[0.222, 0.557]	[-0.141, 0.088]	[0.411, 0.625]
	(-0.040, 0.521)	(-0.769, 0.226)	(0.005, 0.829)	(-0.470, 0.313)	(0.189, 0.884)
Weight on $c \to p$ effect	[0.977, 1.000]	[0.991, 1.000]	[0.947, 1.000]	[0.989, 1.000]	[0.966, 1.000]
Weight on $n \to p$ effect	[0.000, 0.054]	[0.000, 0.037]	[0.000, 0.086]	[0.000, 0.043]	[0.000, 0.074]
Weight on $n \to c$ effect	[0.065, 0.088]	[0.030, 0.045]	[0.131, 0.169]	[0.042, 0.058]	[0.124, 0.152]

This table reports treatment effects of conviction and incarceration, aggregated using the weights from a 2SLS regression with incarceration as the treatment and judge dummies as the instruments. MTRs are approximated by a second-degree polynomial in u_1 and u_2 as specified in Section 5.5. Bounds in square brackets and 95% confidence intervals calculated using Bei (2023) in parentheses.

Table 5: Policy effects after 5 years

	Change in treatment shares			Effects on outcomes	
	n	c	p	N. charges	N. conv.
Panel A: Felony defendants					
Incarceration leniency ($g_2 \downarrow$)	0.000	[0.009, 0.010]	[-0.010, -0.009]	[0.356, 0.660]	[0.256, 0.609]
				(0.248, 0.910)	(0.137, 0.892)
Conviction leniency ($g_1 \downarrow$)	0.010	[-0.010, -0.007]	[-0.003, 0.000]	[-0.032, 0.195]	[-0.013, 0.128]
				(-0.269, 0.361)	(-0.276, 0.406)
No incarceration ($g_2 = 0$)	0.000	0.289	-0.289	[0.085, 0.295]	[0.102, 0.270]
				(-0.001, 0.349)	(0.016, 0.316)
No conviction ($g_1 = 0$)	0.874	-0.585	-0.289	[-9.848, -9.799]	[-4.721, -4.662]
				(-18.950, -0.693)	(-16.024, 6.645)
Homogenize judges $\big(g(z,x) = g(x)\big)$	0.000	0.000	0.000	[-0.201, 0.134]	[-0.204, 0.095]
				(-0.292, 0.320)	(-0.307, 0.294)
Panel B: Misdemeanor defendants					
Incarceration leniency ($g_2 \downarrow$)	0.000	[0.005, 0.010]	[-0.010, -0.005]	[-0.265, 0.190]	[-0.174, 0.383]
				(-0.393, 0.369)	(-0.346, 0.621)
Conviction leniency ($g_1 \downarrow$)	0.010	[-0.010, -0.008]	[-0.002, 0.000]	[-0.112, 0.215]	[-0.449, -0.052]
				(-0.305, 0.382)	(-0.682, 0.140)
No incarceration ($g_2 = 0$)	0.000	0.103	-0.104	[-0.038, 0.030]	[-0.033, 0.047]
				(-0.053, 0.055)	(-0.051, 0.079)
No conviction ($g_1 = 0$)	0.530	-0.427	-0.104	[-0.395, -0.386]	[-0.869, -0.858]
				(-1.027, 0.245)	(-1.528, -0.199)
Homogenize judges $\big(g(z,x) = g(x)\big)$	0.000	0.000	0.000	[-0.024, 0.252]	[-0.041, 0.354]
				(-0.137, 0.336)	(-0.211, 0.468)
Panel C: Never-convicted misdemeanor defendants					
Incarceration leniency ($g_2 \downarrow$)	0.000	[0.005, 0.010]	[-0.010, -0.005]	[-0.193, 0.123]	[-0.186, 0.252]
				(-0.328, 0.314)	(-0.392, 0.530)
Conviction leniency ($g_1 \downarrow$)	0.010	[-0.010, -0.008]	[-0.002, 0.000]	[-0.247, -0.022]	[-0.489, -0.224]
				(-0.420, 0.134)	(-0.697, -0.032)
No incarceration ($g_2 = 0$)	0.000	0.087	-0.087	[-0.021, 0.010]	[-0.018, 0.025]
				(-0.035, 0.031)	(-0.036, 0.049)
No conviction ($g_1 = 0$)	0.528	-0.441	-0.087	[-0.447, -0.436]	[-0.731, -0.715]
				(-1.030, 0.148)	(-1.323, -0.123)
Homogenize judges $\big(g(z,x) = g(x)\big)$	0.000	0.000	0.000	[-0.008, 0.197]	[-0.054, 0.264]
				(-0.151, 0.283)	(-0.213, 0.356)

Table reports the effects of a number of policy changes on recidivism. We analyze marginal changes, which shift judges' thresholds g by 0.01, as well as global changes. The change in treatment shares is the change from the given policy. The change in outcomes is rescaled by the number of defendants whose treatment is affected by the policy change for the marginal changes to assist in readability. Bounds are in square brackets and the outer edges of 95% confidence intervals in parentheses are calculated using Bei (2023).

References

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

Agan, Amanda Y, Andrew Garin, Dmitri K Koustas, Alexandre Mas, and Crystal Yang. 2023. “Labor Market Impacts of Reducing Felony Convictions.” National Bureau of Economic Research.

Agan, Amanda Y, Jennifer L Doleac, and Anna Harvey. 2022. “Misdemeanor prosecution.” National Bureau of Economic Research.

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

Ali, Mir M, NN Mikhail, and M Safiul Haq. 1978. “A class of bivariate distributions including the bivariate logistic.” *Journal of multivariate analysis*, 8(3): 405–412.

Angrist, Joshua D, and Guido W Imbens. 1995. “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.

Arteaga, Carolina. 2021. “Parental Incarceration and Children’s Educational Attainment.” *The Review of Economics and Statistics*, 1–45.

Autor, David, Nicole Maestas, Kathleen Mullen, and Alexander Strand. 2011. “Does delay cause decay? The effect of administrative decision time on the labor force participation and earnings of disability applicants.” *Ann Arbor, MI: University of Michigan Retirement Research Center.*

Bebchuk, Lucian Arye. 1984. “Litigation and settlement under imperfect information.” *The RAND Journal of Economics*, 404–415.

Bei, Xinyue. 2023. “Inference on Union Bounds with Applications to DiD, RDD, Bunching, and Structural Counterfactuals.”

Bhuller, Manudeep, and Henrik Sigstad. 2022. “2SLS with multiple treatments.” *arXiv preprint arXiv:2205.07836.*

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad. 2020. “Incarceration, recidivism, and employment.” *Journal of Political Economy*, 128(4): 1269–1324.

BJS. 2020. “Prisoners in 2019.” *Bureau of Justice Statistics.*

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky. 2022. “When is TSLS actually late?” National Bureau of Economic Research.

Brinch, Christian N, Magne Mogstad, and Matthew Wiswall. 2017. “Beyond LATE with a discrete instrument.” *Journal of Political Economy*, 125(4): 985–1039.

Chien, Colleen. 2020. “America’s paper prisons: The second chance gap.” *Mich. L. Rev.*, 119: 519.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2016. “From LATE to MTE: Alternative methods for the evaluation of policy interventions.” *Labour Economics*, 41: 47–60.

Dobbie, Will, Jacob Goldin, and Crystal S Yang. 2018. “The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges." *American Economic Review*, 108(2): 201–240.

Doyle, Joseph J, John A Graves, Jonathan Gruber, and Samuel A Kleiner. 2015. "Measuring returns to hospital care: Evidence from ambulance referral patterns." *Journal of Political Economy*, 123(1): 170–214.

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

Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2023. "Judging judge fixed effects." *American Economic Review*, 113(1): 253–277.

Garin, Andrew, Dmitri Koustas, Carl McPherson, Samuel Norris, Matthew Pecenco, Evan K Rose, Yotam Shem-Tov, and Jeffrey Weaver. 2023. "The Impact of Incarceration on Employment, Earnings, and Tax Filing." *University of Chicago, Becker Friedman Institute for Economics Working Paper*, , (2023-108).

Green, Donald P, and Daniel Winik. 2010. "Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders." *Criminology*, 48(2): 357–387.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The heavy costs of high bail: Evidence from judge randomization." *The Journal of Legal Studies*, 45(2): 471–505.

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

Heaton, Paul, Sandra Mayson, and Megan Stevenson. 2017. "The downstream consequences of misdemeanor pretrial detention." *Stan. L. Rev.*, 69: 711.

Heckman, James J, and Edward J Vytlacil. 1999. "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the national Academy of Sciences*, 96(8): 4730–4734.

Heckman, James J, and Edward Vytlacil. 2005. "Structural equations, treatment effects, and econometric policy evaluation 1." *Econometrica*, 73(3): 669–738.

Heckman, James J, and Rodrigo Pinto. 2018. "Unordered monotonicity." *Econometrica*, 86(1): 1–35.

Heckman, James J, and Sergio Urzua. 2010. "Comparing IV with structural models: What simple IV can and cannot identify." *Journal of Econometrics*, 156(1): 27–37.

Heckman, James J, Sergio Urzua, and Edward Vytlacil. 2006. "Understanding instrumental variables in models with essential heterogeneity." *The review of economics and statistics*, 88(3): 389–432.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. 2008. "Instrumental Variables in Models with Multiple Outcomes: the General Unordered Case." *Annales d'Économie et de Statistique*, , (91/92): 151–174.

Hull, Peter. 2020. "Estimating hospital quality with quasi-experimental data." *Available at SSRN 3118358*.

Humphries, John Eric, Aurelie Ouss, Kamelia Stavreva, Megan T Stevenson, and Winnie van Dijk. 2023. "Conviction, Incarceration, and Recidivism: Understanding the

Revolving Door."

Huttunen, Kristiina, Martti Kaila, and Emily Nix. 2021. "The Punishment Ladder: Estimating the Impact of Different Punishments on Defendant Outcomes."

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

Jordan, Andrew, Ezra Karger, and Derek Neal. 2023. "Heterogeneous impacts of sentencing decisions." National Bureau of Economic Research.

Kamat, Vishal, Samuel Norris, and Matthew Pecenco. 2023. "Identification in Multiple Treatment Models under Discrete Variation."

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad. 2016. "Field of study, earnings, and self-selection." *The Quarterly Journal of Economics*, 131(3): 1057–1111.

Kline, Patrick, and Christopher R Walters. 2016. "Evaluating public programs with close substitutes: The case of Head Start." *The Quarterly Journal of Economics*, 131(4): 1795–1848.

Kling, Jeffrey R. 2006. "Incarceration length, employment, and earnings." *American Economic Review*, 96(3): 863–876.

Kuziemko, Ilyana. 2012. "How should inmates be released from prison? An assessment of parole versus fixed-sentence regimes." *The Quarterly Journal of Economics*, 128(1): 371–424.

Landes, William M. 1971. "An economic analysis of the courts." *The Journal of Law and Economics*, 14(1): 61–107.

Lee, Sokbae, and Bernard Salanié. 2018. "Identifying effects of multivalued treatments." *Econometrica*, 86(6): 1939–1963.

Loeffler, Charles E. 2013. "Does imprisonment alter the life course? Evidence on crime and employment from a natural experiment." *Criminology*, 51(1): 137–166.

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

Mai, Chris, and Ram Subramanian. 2017. "The price of prisons: Examining state spending trends, 2010-2015." *Vera Institute of Justice*, 7–8.

Miller, Ted R, Mark A Cohen, David I Swedler, Bina Ali, and Delia V Hendrie. 2021. "Incidence and costs of personal and property crimes in the USA, 2017." *Journal of Benefit-Cost Analysis*, 12(1): 24–54.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R Walters. 2021. "The causal interpretation of two-stage least squares with multiple instrumental variables." *American Economic Review*, 111(11): 3663–98.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. "Using instrumental variables for inference about policy relevant treatment parameters." *Econometrica*, 86(5): 1589–1619.

Mountjoy, Jack. 2022. "Community colleges and upward mobility." *Available at SSRN 3373801.*

Mueller-Smith, Michael. 2015. "The criminal and labor market impacts of incarceration." *Unpublished Working Paper*, 18.

Mueller-Smith, Michael, and Kevin Schnepel. 2021. "Diversion in the criminal justice system." *The Review of Economic Studies*, 88(2): 883–936.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver. 2021. "The effects of parental and sibling incarceration: Evidence from ohio." *American Economic Review*, 111(9): 2926–63.

Ohio Criminal Sentencing Commission. 2021. "The Ohio Sentencing Data Platform: Goals and Uses."

Ostrom, Brian J, Lydia E Hamblin, Richard Y Schauffler, and Nial Raaen. 2020. "Timely Justice in Criminal Cases: What the Data Tells Us." *National Center for State Courts. https://www. ncsc. org/__data/assets/pdf_file/0019/53218/Timely-Justice-in-Criminal-Case s-What-the-Data-Tells-Us. pdf.*

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

Priest, George L, and Benjamin Klein. 1984. "The selection of disputes for litigation." *The journal of legal studies*, 13(1): 1–55.

Rivera, Roman. 2023. "Release, Detain, or Surveil? The Effect of Electronic Monitoring on Defendant Outcomes." *Unpublished working paper.*

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

Shen, Yinzhi, Shawn D Bushway, Lucy C Sorensen, and Herbert L Smith. 2020. "Locking up my generation: Cohort differences in prison spells over the life course." *Criminology*, 58(4): 645–677.

Silveira, Bernardo S. 2017. "Bargaining with asymmetric information: An empirical study of plea negotiations." *Econometrica*, 85(2): 419–452.

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

Vytlacil, EJ. 2006. "Ordered discrete choice selection models: Equivalence, nonequivalence, and representation results." *Review of Economics and Statistics*, 88(3): 578–581.

Zeng, Z. 2020. "Jail inmates in 2018." *US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC, available at: www. bjs. gov/content/pub/pdf/ji18. pdf (accessed 15 June 2020).*

Appendix

A1 Implications of plea bargaining for choice model

An important aspect of our empirical setting is the prevalence of plea bargaining. As we discuss in Section 2.2, many cases end before a trial with the defendant pleading guilty to at least some charges. In such cases, the prosecutor and defense attorney enter a joint recommendation to the judge, who determines an appropriate sentence. In this section, we theoretically analyze the effect of plea bargaining on our choice model. We assume that our choice model is satisfied before the introduction of plea bargaining and then study whether the introduction of pleas derails the model. Interestingly, we find that although pretrial bargaining changes the case outcome for some individuals, the same monotonicity conditions continue to characterize the decisions.

We adopt a Nash bargaining framework with full information. Institutionally, plea bargaining can occur at any time prior to trial and often occurs shortly before a scheduled trial date. Consequently, the full information framework is a realistic depiction of the state of information after pre-trial discovery, which requires the prosecution to provide the defense with almost all of their collected materials, including all materials they will use in the trial and any exculpatory evidence.[1]

Nash bargaining model

We suppose that the court must decide between three outcomes, $i \in \{0, 1, 2\}$, which correspond to case dismissal, conviction without incarceration, and incarceration. For case outcome i, the defense attorney receives utility α_i and the prosecutor receives utility β_i. We assume that the punishment becomes more severe in i, and so $\alpha_0 \geq \alpha_1 \geq \alpha_2$ and $\beta_0 \leq \beta_1 \leq \beta_2$.

We assume that the defense attorney and prosecutor Nash bargain over the outcome, with the fallback option of having a trial occurring under bargaining breakdown. In case of a trial, the judge will decide the outcome and the defense attorney and prosecutor will pay a fixed cost $\alpha^* \geq 0$ and $\beta^* \geq 0$, respectively.

This is a full information setup, so the participants are aware the judge would choose option $k(z)$ if negotiations break down. The participants will then agree on the following outcome:

$$\underset{i}{\operatorname{argmax}} \;\; V_i\big(k(z)\big), \qquad V_i\big(k(z)\big) = \big(\alpha_i - (\alpha_{k(z)} - \alpha^*)\big)\big(\beta_i - (\beta_{k(z)} - \beta^*)\big)$$

where $V_i\big(k(z)\big)$ is the Nash product under judge z.

Plea bargaining does not cause monotonicity violations

LM requires that judges behave monotonically when deciding between conviction (either c or p) and n, and also monotonically between p and c when deciding between those two choices.

[1] An important source of asymmetric information can be the whereabouts or status of witnesses. As this too typically becomes known prior to the trial, this information also becomes reflected in plea bargains.

We begin by assuming the judges behave according to LM, and then show that the decisions after plea bargaining will continue to obey monotonicity. To emphasize that the defendants and prosecutors preferences are ordered, we index treatment with $\{0, 1, 2\}$ rather than $\{n, c, p\}$ as in the main paper.

We first consider the requirement that when choosing between options 2 and 1, we have that

$$D_2^*(z) \geq D_2^*(z') \text{ or } D_2^*(z) \leq D_2^*(z') \ .$$

where $D_2^*(z)$ is a dummy variable indicating that judges z chooses treatment 2 when he is choosing between options 1 and 2. Without loss of generality, we assume that $D_2^*(z) \geq D_2^*(z')$. Then, for some new set of decisions under plea bargaining, which we denote $\widetilde{D}$, monotonicity is violated if there is a case where

$$D_2^*(z) > D_2^*(z') \tag{A1}$$
$$\widetilde{D}_2^*(z) < \widetilde{D}_2^*(z'). \tag{A2}$$

(A1) implies that $k(z) = 2$ and $k(z') = 1$ and (A2) implies that $V_2\big(k(z)\big) < V_1\big(k(z)\big)$ and $V_2\big(k(z')\big) > V_1\big(k(z')\big)$. Beginning with the first inequality, and using the Nash products resulting from (A1), we find

$$\begin{aligned}
V_2\big(k(z)\big) <& V_1\big(k(z)\big) \\
\alpha^*\beta^* <& \alpha^*\beta^* + (\alpha_1 - \alpha_2)\beta^* + (\beta_1 - \beta_2)\alpha^* + (\alpha_1 - \alpha_2)(\beta_1 - \beta_2) \\
0 <& (\alpha_1 - \alpha_2)\beta^* + (\beta_1 - \beta_2)\alpha^* + (\alpha_1 - \alpha_2)(\beta_1 - \beta_2) \\
2(\alpha_2 - \alpha_1)(\beta_2 - \beta_1) >& (\alpha_2 - \alpha_1)\beta^* + (\beta_2 - \beta_1)\alpha^* + (\alpha_2 - \alpha_1)(\beta_2 - \beta_1) \\
0 >& (\alpha_2 - \alpha_1)\beta^* + (\beta_2 - \beta_1)\alpha^* + (\alpha_2 - \alpha_1)(\beta_2 - \beta_1) \qquad \text{(A3)} \\
\alpha^*\beta^* >& \alpha^*\beta^* + (\alpha_2 - \alpha_1)\beta^* + (\beta_2 - \beta_1)\alpha^* + (\alpha_2 - \alpha_1)(\beta_2 - \beta_1) \\
V_1\big(k(z')\big) >& V_2\big(k(z')\big)
\end{aligned}$$

where (A3) follows because α_i is monotonically decreasing and β_i is monotonically increasing in i, so $2(\alpha_2 - \alpha_1)(\beta_2 - \beta_1) < 0$. The last line directly contradicts (A2). This shows that when judges are choosing whether to incarcerate (conditional on conviction), plea bargaining is not enough to overturn monotonicity. When plea bargaining moves the likelihood of defendants being treated in a particular direction, it does so monotonically for all judges.

Next, we show that if $D_{12}(z) \geq D_{12}(z')$, then $\widetilde{D}_{12}(z) \geq \widetilde{D}_{12}(z')$. Suppose not. Then, $D_{12}(z) > D_{12}(z')$ and $\widetilde{D}_{12}(z) < \widetilde{D}_{12}(z')$. We consider the case $k(z) = 2$; the proof for when $k(z) = 1$ is almost identical.

When $k(z) = 2$ and $k(z') = 0$, the Nash products are

$$V_2\big((k(z)\big) = \alpha^*\beta^* \tag{A4}$$

$$V_1\big((k(z)\big) = \alpha^*\beta^* + (\alpha_1 - \alpha_2)\beta^* + (\beta_1 - \beta_2)\alpha^* + (\alpha_1 - \alpha_2)(\beta_1 - \beta_2) \tag{A5}$$

$$V_0\big((k(z)\big) = \alpha^*\beta^* + (\alpha_0 - \alpha_2)\beta^* + (\beta_0 - \beta_2)\alpha^* + (\alpha_0 - \alpha_2)(\beta_0 - \beta_2) \tag{A6}$$

$$V_2\big((k(z')\big) = \alpha^*\beta^* + (\alpha_2 - \alpha_0)\beta^* + (\beta_2 - \beta_0)\alpha^* + (\alpha_2 - \alpha_0)(\beta_2 - \beta_0) \tag{A7}$$

$$V_1\big((k(z')\big) = \alpha^*\beta^* + (\alpha_1 - \alpha_0)\beta^* + (\beta_1 - \beta_0)\alpha^* + (\alpha_1 - \alpha_0)(\beta_1 - \beta_0) \tag{A8}$$

$$V_0\big((k(z')\big) = \alpha^*\beta^* \tag{A9}$$

Monotonicity will be violated if $\widetilde{D}(z) = 0$ and either $\widetilde{D}(z') = 1$ or $\widetilde{D}(z') = 2$. If $\widetilde{D}(z') = 2$, then we find a contradiction analogously to the two-option case. If $\widetilde{D}(z') = 1$, then we have that $V_1\big(z(k')\big) > V_0\big(z(k')\big)$, implying

$$(\alpha_1 - \alpha_0)\beta^* + (\beta_1 - \beta_0)\alpha^* + (\alpha_1 - \alpha_0)(\beta_1 - \beta_0) > 0 \tag{A10}$$

$$(\alpha_0 - \alpha_1)\beta^* + (\beta_0 - \beta_1)\alpha^* < (\alpha_1 - \alpha_0)(\beta_1 - \beta_0) \tag{A11}$$

$$(\alpha_0 - \alpha_1)\beta^* + (\beta_0 - \beta_1)\alpha^* < 0 \tag{A12}$$

where the third line follows because α_i is increasing in i and β_i is decreasing.

Similarly, $\widetilde{D}(z) = 0$ so $V_0\big(z(k)\big) > V_1\big(z(k)\big)$ and

$$(\alpha_0 - \alpha_1)\beta^* + (\beta_0 - \beta_1)\alpha^* + (\alpha_0 - \alpha_2)(\beta_0 - \beta_2) - (\alpha_1 - \alpha_2)(\beta_1 - \beta_2) > 0$$
$$(\alpha_0 - \alpha_1)\beta^* + (\beta_0 - \beta_1)\alpha^* + (\alpha_0 - \alpha_1)(\beta_0 - \beta_1) + (\alpha_0 - \alpha_1)(\beta_1 - \beta_2) + (\alpha_1 - \alpha_2)(\beta_0 - \beta_1) > 0$$

Given that each of the interaction terms is negative (because α_i and β_i are ordered in opposite directions), by combining with (A12) we arrive at a contradiction.

To conclude, if judge decisions obey LM without plea bargaining, in this simple model they continue to obey LM with plea bargaining. While more complicated models that incorporate private information (e.g., Silveira (2017)) can produce violations of latent monotonicity, this section shows that when private information is limited, so will violations of LM.

A2 Interpreting the 2SLS estimand

In this section we provide a decomposition of the 2SLS estimand into its constituent effects. To maintain clarity, we initially abstract away from covariates. However, after deriving the result, we write out our estimating equation in the index representation and include covariates as in our baseline specification.

The building blocks of our analysis are the compliance groups defined by the intersection of their treatment assignment under each judge $\{\bigcap_{z\in\mathcal{Z}} D(z)\}$. For compliance group ℓ define the measure of that group as π_ℓ, and the average treatment effect of s versus t as $\Delta^\ell_{st} = E[Y(s) - Y(t) \mid C_\ell]$ where C_ℓ denotes membership in group ℓ. Using $D_{\ell j}$ as shorthand for $D(z_j)$ for compliance group ℓ, we decompose the binary incarceration 2SLS estimand as

$$
\begin{aligned}
\beta^{2\text{SLS}}_{incar} &= \sum_{j=1}^{J} \lambda_j \frac{E[Y|z_j] - E[Y|z_{j-1}]}{E[D_p|z_j] - E[D_p|z_{j-1}]} \qquad \text{(A13)}\\
&= \sum_{j=1}^{J} \frac{\lambda_j}{E[D_p|z_j] - E[D_p|z_{j-1}]} \Big(\sum_\ell \Delta^\ell_{pc} \mathbb{1}[D_{\ell j}{=}p, D_{\ell,j-1}{=}c]\pi_\ell + \sum_\ell \Delta^\ell_{pn} \mathbb{1}[D_{\ell j}{=}p, D_{\ell,j-1}{=}n]\pi_\ell + \\
&\qquad \sum_\ell \Delta^\ell_{cp} \mathbb{1}[D_{\ell j}{=}c, D_{\ell,j-1}{=}p]\pi_\ell + \sum_\ell \Delta^\ell_{cn} \mathbb{1}[D_{\ell j}{=}c, D_{\ell,j-1}{=}n]\pi_\ell + \\
&\qquad \sum_\ell \Delta^\ell_{np} \mathbb{1}[D_{\ell j}{=}n, D_{\ell,j-1}{=}p]\pi_\ell + \sum_\ell \Delta^\ell_{nc} \mathbb{1}[D_{\ell j}{=}n, D_{\ell,j-1}{=}c]\pi_\ell \Big) \\
&= \sum_\ell \phi^\ell_{pc}\Delta^\ell_{pc} + \phi^\ell_{pn}\Delta^\ell_{pn} + \phi^\ell_{cp}\Delta^\ell_{cp} + \phi^\ell_{cn}\Delta^\ell_{cn} + \phi^\ell_{np}\Delta^\ell_{np} + \phi^\ell_{nc}\Delta^\ell_{nc}
\end{aligned}
$$

where λ_j is the classic 2SLS weight from Imbens and Angrist (1994) arising from instrumenting for incarceration (treatment p) with judge indicators, and

$$
\begin{aligned}
\phi^\ell_{st} &= (\widetilde{\phi}^\ell_{st} - \widetilde{\phi}^\ell_{ts})\mathbb{1}[\widetilde{\phi}^\ell_{st} - \widetilde{\phi}^\ell_{ts} > 0] \\
\widetilde{\phi}^\ell_{st} &= \sum_{j=1}^{J} \frac{\lambda_j}{E[D_p|z_j] - E[D_p|z_{j-1}]} \mathbb{1}[D_{\ell j}{=}s, D_{\ell,j-1}{=}t]\pi_\ell
\end{aligned}
$$

This decomposition of $\beta^{2\text{SLS}}_{incar}$ exploits the fact that for each pair of treatments s and t, the 2SLS-weighted judge assignment induces a compliance group either from s to n or vice versa. By construction, ϕ^ℓ_{st} is always weakly positive and, when it is strictly positive, represents the weight on the treatment effect for individuals who move from treatment t to s when they are assigned to the j^{th} rather than the $j-1^{\text{th}}$ most severe judge.

We then define the average treatment effect for individuals induced from treatment t to s as

$$
\Delta_{st} = \frac{\sum_\ell \phi^\ell_{st}\Delta^\ell_{st}}{\sum_\ell \phi^\ell_{st}} \qquad \text{(A14)}
$$

and the weight as $\phi_{st} = \sum_\ell \phi^\ell_{st}$. This lets us rewrite the 2SLS estimand, which is discussed in the main text as (6), as

$$\beta^{\text{2SLS}}_{incar} = \phi_{pc}\Delta_{pc} + \phi_{pn}\Delta_{pn} + \phi_{cp}\Delta_{cp} + \phi_{np}\Delta_{np} + \phi_{cn}\Delta_{cn} + \phi_{nc}\Delta_{nc} \tag{A15}$$

Representation in index form

In this subsection we re-derive (A13) using index notation and accounting for covariates. This provides the components in (11).

In the case where instruments are saturated in covariates, we can decompose the 2SLS estimand as a weighted combination of covariate-cell-specific LATEs. Letting j index judges and ordering them up to $J(x)$ in each cell c by propensity to incarcerate, these weights are equal to

$$w_x = \frac{P[x]\bigg(\sum_{j=0}^{J(x)} P[z_j \mid x]\Big(E[D_p|z_j,x] - E[D_p|x]\Big)^2\bigg)}{\sum_{x'} P[x']\bigg(\sum_{j=0}^{J(x')} P[z_j \mid x']\Big(E[D_p|z_j,x] - E[D_p|x']\Big)^2\bigg)}$$

Using $\Delta_{st}(u,x) \equiv m_s(u,x) - m_t(u,x)$ to denote the effect of s relative to t for a u-indexed individual in covariate cell x, and $D_{j,x}(u)$ to denote the treatment decision of judge j in cell x, we can then decompose the 2SLS estimand as

$$\begin{aligned}
\beta^{\text{2SLS}}_{incar} &= \sum_x w_x \sum_{j=1}^{J(x)} \lambda_{jx} \frac{E[Y|z_j,x] - E[Y|z_{j-1},x]}{E[D_p|z_j,x] - E[D_p|z_{j-1},x]} && \text{(A16)}\\
&= \sum_x w_x \sum_{j=1}^{J} \frac{\lambda_{jx}}{E[D_p|z_j,x] - E[D_p|z_{j-1},x]} \times \\
&\quad \bigg(\int \Delta_{pc}(u,x)\mathbb{1}[D_{jx}(u){=}p, D_{j-1,x}(u){=}c]f(u) + \Delta_{pn}(u,x)\mathbb{1}[D_{jx}(u){=}p, D_{j-1,x}(u){=}n]f(u) + \\
&\quad \Delta_{cp}(u,x)\mathbb{1}[D_{jx}(u){=}c, D_{j-1,x}(u){=}p]f(u) + \Delta_{cn}(u,x)\mathbb{1}[D_{jx}(u){=}c, D_{j-1,x}(u){=}n]f(u) + \\
&\quad \Delta_{np}(u,x)\mathbb{1}[D_{jx}(u){=}n, D_{j-1,x}(u){=}p]f(u) + \Delta_{nc}(u,x)\mathbb{1}[D_{jx}(u){=}n, D_{j-1,x}(u){=}c]f(u)\ du\bigg) \\
&= \sum_x w_x \int \phi_{pc}(u,x)\Delta_{pc}(u,x) + \phi_{pn}(u,x)\Delta_{pn}(u,x) + \phi_{cp}(u,x)\Delta_{cp}(u,x)\ + && \text{(A17)}\\
&\qquad \phi_{cn}(u,x)\Delta_{cn}(u,x) + \phi_{np}(u,x)\Delta_{np}(u,x) + \phi_{nc}(u,x)\Delta_{nc}(u,x)\ du
\end{aligned}$$

where λ_{jx} are the within-x 2SLS weights and analogously to (A13),

$$\begin{aligned}
\phi_{st}(u,x) &= (\widetilde{\phi}_{st}(u,x) - \widetilde{\phi}_{ts}(u,x))\mathbb{1}[\widetilde{\phi}_{st}(u,x) - \widetilde{\phi}_{ts}(u,x) > 0] \\
\widetilde{\phi}_{st}(u,x) &= \sum_{j=1}^{J(x)} \frac{\lambda_{jx}}{E[D_p|z_j,x] - E[D_p|z_{j-1},x]} \mathbb{1}[D_{jx}(u){=}s, D_{j-1,x}(u){=}t]f(u)
\end{aligned}$$

We can then define the weights and margin-specific treatment effects as

$$
\begin{aligned}
\Delta_{s\to t} &= \sum_x w_x \int [m_t(u,x) - m_s(u,x)] w_{s\to t}(u,x)\ du \\
w_{s\to t}(u,x) &= \frac{\phi_{ts}(u,x)}{\sum_x w_x \int \phi_{ts}(v,x)\ dv}
\end{aligned}
$$

where the treatment effects reflect the weighted effect of compliers who move from s to t as a result of judge assignment. Since the compliance groups depend on the distribution of u, note that the weights and treatment effects are themselves functions of F. See (6) and (11), where we use these estimates to construct our target parameters.

A3 Instrumenting for multiple treatments at once

In this section, we discuss the interpretation of the 2SLS estimand when the researcher simultaneously instruments for both conviction and incarceration. We show that the estimands reflect a combination of the effects of different treatments and for different complier groups, defying a causal interpretation even under restrictive patterns of compliance. We conclude that the results of these regressions should be interpreted as causal effects only under the assumption of constant effects.

We consider the following 2SLS specification:

$$Y_i = \beta_0 + \beta_c \mathbb{1}[D_i{=}c] + \beta_p \mathbb{1}[D_i{=}p] + \varepsilon_i \tag{A18}$$

$$\mathbb{1}[D_i{=}c] = \alpha_j^c + e_i \tag{A19}$$

$$\mathbb{1}[D_i{=}p] = \alpha_j^p + u_i \tag{A20}$$

where β_c and β_p are the coefficients of interest and α represents the judge indicators used as the instruments.

By Frisch-Waugh-Lovell, these estimands can be decomposed as

$$\beta_c = \frac{\frac{\text{Cov}(Y,P_c)}{\text{Var}(P_c)} - \frac{\text{Cov}(P_c,P_p)}{\text{Var}(P_c)}\frac{\text{Cov}(Y,P_p)}{\text{Var}(P_p)}}{1-\rho_{cp}^2}$$

$$\beta_p = \frac{\frac{\text{Cov}(Y,P_p)}{\text{Var}(P_p)} - \frac{\text{Cov}(P_c,P_p)}{\text{Var}(P_p)}\frac{\text{Cov}(Y,P_c)}{\text{Var}(P_c)}}{1-\rho_{cp}^2}$$

where $P_d = E[D{=}d|Z=j]$ for each judge j and $\rho_{cp} \equiv \text{Cov}(P_c, P_p)/\sqrt{\text{Var}(P_c)\text{Var}(P_p)}$ is the correlation between P_c and P_p.

The expression for β_c reveals that the coefficient on the conviction dummy in (A18) is equal to the coefficient from a 2SLS regression of the outcome on instrumented conviction, minus the coefficient from a 2SLS regression of the outcome on instrumented incarceration multiplied by the effect of judge-instrumented conviction on incarceration, all rescaled by a term involving the correlation between the two treatment propensities.

Analogously to the single-treatment case in Appendix A2, these expressions can be decomposed into a weighted combination of treatment effects for the compliers corresponding to compliance group ℓ across each of the three treatments. In particular, it is easy to use the arguments in Appendix A2 to show that

$$\beta_c = \sum_\ell \phi_{pc}^{\ell c}\Delta_{pc}^\ell + \phi_{pn}^{\ell c}\Delta_{pn}^\ell + \phi_{cp}^{\ell c}\Delta_{cp}^\ell + \phi_{cn}^{\ell c}\Delta_{cn}^\ell + \phi_{np}^{\ell c}\Delta_{np}^\ell + \phi_{nc}^{\ell c}\Delta_{nc}^\ell \tag{A21}$$

where

$$\phi^{\ell c}_{st} = \max\Big(0, (\widetilde{\phi}^{\ell c}_{st} - \psi_c \widetilde{\phi}^{\ell p}_{st}) - (\widetilde{\phi}^{\ell c}_{ts} - \psi_c \widetilde{\phi}^{\ell p}_{ts})\Big)$$

$$\widetilde{\phi}^{\ell d}_{st} = \sum_{j_d=1}^{J} \frac{\lambda^d_j}{E[D_d|z_{j_d}] - E[D_d|z_{j_d-1}]} \mathbb{1}[D_{\ell j_d}{=}s, D_{\ell,j_d-1}{=}t]\pi_\ell$$

and where λ^d_j is the classic 2SLS weight from Imbens and Angrist (1994) arising from instrumenting for being in treatment d with judge indicators. $\psi_c = \frac{\text{Cov}(P_c,P_p)}{\text{Var}(P_c)}$ is the coefficient from a regression of judges' p treatment share on their c treatment share. We sub-index the judge indices with d to denote that the ordering is treatment-specific. The expression for β_p is the same as that for β_c, but using $\phi^{\ell p}$ weights.

The weights in (A21) depend on the underlying compliance patterns and are positive by construction, but do not necessarily sum to one. This means it is in general not possible to interpret β_c and β_p as positively-weighted combinations of treatment effects because of compliers moving in opposite directions across treatments.

Recent research has investigated conditions under which 2SLS will deliver interpretable treatments effects. Bhuller and Sigstad (2022) show (Proposition B.8) that one such condition is the combination of both strict Imbens and Angrist (1994) monotonicity for each treatment and judge pair, as well as mutual linearity of the judge propensities for treatments c and p in each other. These conditions are strict, and not satisfied by typical models of judge decision-making. In particular, our baseline assumption LM does not satisfy the monotonicity condition and therefore will not deliver a convex combination of treatment effects.

To highlight the issues with 2SLS, we examine in more detail a single-index model of treatment (Heckman and Vytlacil, 2005; Rivera, 2023). This model is more restrictive than our baseline model, since it is the special case of LM for ρ=1. Nonetheless, it is not restrictive enough for 2SLS to return interpretable treatment effects. To fix ideas, a single-index model assigns treatment in the following way:

Assumption SI *(Single Index) For each judge $z \in \mathcal{Z}$, treatment is determined by*

$$D(z) = \begin{cases} n & \text{if } \; g_1(z) < U \;, \\ c & \text{if } \; g_2(z) < U \leq g_1(z) \;, \\ p & \text{if } \; U \leq g_2(z) \;, \end{cases}$$

where $U \sim U[0,1]$ and $g_2(z) \leq g_1(z)$.

We consider a case with three equally-likely judges who have incarceration and conviction thresholds of $(g_2(z), g_1(z)) \in \{(0.25, 0.30), (0.35, 0.75), (0.40, 0.85)\}$. This generates 5 different compliance groups ($u \in \{(0.25, 0.30], (0.30, 0.35], (0.35, 0.40], (0.40, 0.75], (0.75, 0.85]\}$). Table A8 shows the treatment assignment for each compliance group under each judge. Changes in judge assignment increase the severity of criminal justice contact for each compliance group; moving from judge 0 to 1 moves defendants from n to either c or p, while moving from judge

1 to 2 moves some defendants from c to p, and other defendants from n to c. However, it does not satisfy Imbens and Angrist (1994) monotonicity. For example, changing assignment from judge 1 to judge 2 moves defendants with $u \in (0.78, 0.85]$ into treatment c, but defendants with $u \in (0.35, 0.40]$ *out* of c.

The failure of Imbens and Angrist (1994) monotonicity suggests that 2SLS will not deliver an interpretable treatment effect. To make this point more precise, in Table A9 we calculate the weights on each compliance group in a 2SLS regression of outcomes on conviction and incarceration. The table reveals that the coefficients from (A18) will reflect a combination of treatment effects for different compliance groups. The coefficient on β_p reflects moves from conviction into incarceration (with a weight of 1.4), but also moves from conviction to dismissal (two different compliance groups, for a total weight of 1.6). It will also reflect $n \to c$, $p \to n$, and $p \to c$ effects. Under treatment effect heterogeneity, therefore, β_p will not correspond to an effect of incarceration in any meaningful sense.

In fact, it is difficult to find any situation in which the 2SLS coefficients will represent margin-specific causal effects. Consider the thresholds $(g_2(z), g_1(z)) \in \{(0.0, 0.0), (0.0, 1.0), (0.5, 1.0)\}$, which implies judge 0 convicts no one, judge 1 convicts everyone and incarcerates no one, and judge 2 incarcerates half and convicts the rest. These judges provide possible comparisons that satisfy the unordered partial monotonicity (UPM) assumption in Mountjoy (2022) and therefore identify margin-specific effects; for example, comparing outcomes for judge 1 to 0 identifies the $n \to c$ ATE, and comparing outcomes for judge 2 to judge 1 identifies a $c \to p$ LATE. They also satisfy Imbens and Angrist (1994) monotonicity. However, 2SLS does not recover these effects: the weights on the incarceration term $\phi_{st}^{l_p}$ have equal weight of 0.5 on $n \to c$, $c \to p$, and $n \to p$. As such, the weights do not reflect comparisons of margin-specific causal treatment effects and additionally include compliance for treatment effects, such as dismissal to conviction effects, that do not reflect the treatment of interest. This is due to the nonlinearity of the conditional mean of the p propensity with respect to the c propensity. We conclude that even in propitious conditions, the coefficients from a multiple-treatment 2SLS regression may mix effects across non-target compliance groups. This will result in bias whenever there are heterogeneous treatment effects, although the degree of bias will vary across settings.

A4 Ordered and unordered monotonicity

In this section, we formally compare our choice model to ordered (Angrist and Imbens, 1995), unordered (Heckman and Pinto, 2018), and single-index (Heckman and Vytlacil, 2005; Rivera, 2023) models.

We first reprise some notation. For each $z \in \mathcal{Z}$, let $D_{cp}(z) \equiv 1\{D(z) \in \{c, p\}\}$ denote an indicator for whether the individual is convicted, regardless of being incarcerated or not, and $D_d(z) \equiv 1\{D(z){=}d\}$ denote an indicator for being in treatment d. This lets us succinctly define *joint monotonicity*, which we use as a building block to discuss the other monotonicity assumptions. For clarity, we drop any conditioning on covariates, but each definition can be understood as conditional on x's.

Assumption JM *(Joint Monotonicity) For each $z, z' \in \mathcal{Z}$, we have*

$$\begin{aligned} D_{cp}(z) \geq D_{cp}(z') \quad &\textit{or} \quad D_{cp}(z) \leq D_{cp}(z') \ , \\ D_p(z) \geq D_p(z') \quad &\textit{or} \quad D_p(z) \leq D_p(z') \ . \end{aligned}$$

This assumption presumes that judges can be ordered with respect to their decision to convict (whether or not they incarcerate), and separately by their decision to incarcerate. It is weaker than both ordered and unordered monotonicity. Specifically, ordered monotonicity (Angrist and Imbens, 1995) assumes that

Assumption OM *(Ordered Monotonicity) For each $z, z' \in \mathcal{Z}$, we have*

$$\begin{aligned} D_{cp}(z) \geq D_{cp}(z') \quad \textit{and} \quad D_p(z) \geq D_p(z') \ , \quad \textit{or} \\ D_{cp}(z) \leq D_{cp}(z') \quad \textit{and} \quad D_p(z) \leq D_p(z') \ . \end{aligned}$$

while unordered monotonicity (Heckman and Pinto, 2018) imposes that

Assumption UM *(Unordered Monotonicity) For each $z, z' \in \mathcal{Z}$, we have that Assumption JM is satisfied and*

$$D_c(z) \geq D_c(z') \quad \textit{or} \quad D_c(z) \leq D_c(z') \ .$$

While these assumptions may be appropriate for some settings, in many examiner designs they may impose unrealistically strong restrictions on treatment assignment. To better see the implications of each of these assumptions, Table A7 displays the response types $(D(z), D(z'))$ between a pair of judges $z, z' \in \mathcal{Z}$. We focus here on the case where $D_{cp}(z) \geq D_{cp}(z')$ as all the assumptions impose it without loss of generality.

The first row of Table A7 reveals that OM rules out any defier types—if one judge is more severe in terms of conviction, then each defendant must be more likely to be incarcerated by her. This is particularly problematic in our setting because judges are instructed to treat the conviction decision $D_{cp}(z)$ as separate from the sentencing decision (incarceration conditional on being convicted, $D_p(z)$), so it seems likely that judges might differ in their severity across those two margins.

An alternative to ordered monotonicity is UM. However, the second set of rows in Table A7 reveals that UM also rules out some natural response patterns. These restrictions arise because UM disallows two-way flows into and out of a treatment to ensure identification of the complier shares. Two-way flows with respect to incarceration—i.e., pairs of judges with both $n \rightarrow p$ types and $p \rightarrow c$ types—are particularly natural given that judges might have different orderings of standards for conviction versus incarceration. UM also precludes two-way flows with respect to conviction: a pair of judges cannot have both $n \rightarrow c$ and $c \rightarrow p$ compliers. This substitution pattern would be expected if the judges behave like the treatments can be ordered. By disallowing two-way flows, therefore, UM disallows at least some of the substitution patterns that are likely to occur in our setting.

A4.1 Single-index choice model

One parsimonious alternative to OM and UM is to require that treatment is determined by a single unobservable index (Heckman and Vytlacil, 2005; Rivera, 2023). A key advantage of this model is that it allows the researcher to estimate marginal treatment effects (MTEs) along this single dimension. However, we show in this section that the single index model rules out certain compliance patterns that are important in our setting. Furthermore, we provide a new result that demonstrates that the single-index model is closely related to both OM and UM; if some of the judges satisfy a particular condition relating to the shares of defendants assigned to each treatment, then all three models are identical. We take this as further evidence that OM and UM might be inappropriate for use in examiner designs.

We begin by defining treatment assignment in the single-index model:

Assumption SI *(Single Index) For each judge $z \in \mathcal{Z}$, treatment is determined by*

$$D(z) = \begin{cases} n & \text{if } \; g_1(z) < U \; , \\ c & \text{if } \; g_2(z) < U \leq g_1(z) \; , \\ p & \text{if } \; U \leq g_2(z) \; , \end{cases}$$

where $U \sim U[0,1]$ and $g_2(z) \leq g_1(z)$.

In the single-index model, defendants may be marginal between n and c or between c and p. However, they can only be marginal between n and p if the judge does not assign any defendants to c. In practice, since we don't observe any judges who don't assign anyone to c, this amounts to ruling out situations where judges are marginal between finding a defendant not guilty and incarcerating them. This, in turn, is at odds with accommodating defendants who face severe charges (and so would be incarcerated if convicted) but are marginal on whether they will be convicted.

We display the full set of allowable response types under SI in Table A7. The table reveals that SI also rules out two-way flows in and out of incarceration, another key substitution pattern that we seek to accommodate. We conclude that the single-index model is unlikely to be appropriate in our setting.

While SI may appear to be more restrictive than both OM and UM, there are in fact deep underlying similarities between the models. To demonstrate this, we focus on JM, which is weaker than both OM and UM. Building on Vytlacil (2002, 2006), we provide an index characterization of JM, then demonstrate an important condition under which it is equivalent to SI.

Proposition A1 *Assumption JM is equivalent to*

$$D(z) = \begin{cases} n & \text{if } \; U_1 > g_1(z) \; , \\ c & \text{if } \; U_1 \leq g_1(z), U_2 > g_2(z) \; , \\ p & \text{if } \; U_1 \leq g_1(z), U_2 \leq g_2(z) \; , \end{cases} \tag{A22}$$

for each $z \in \mathcal{Z}$, where $U_1, U_2 \sim U[0,1]$ and $g_1(z) \geq g_2(z)$, and

$$P[U_1 > g_1(z), U_2 \leq g_2(z)] = 0 \tag{A23}$$

$$P[U_1 \leq g_1(z), U_2 > g_2(z)] = g_2(z) - g_1(z) \tag{A24}$$

Furthermore, if for every $t \in [0,1]$ there exists $z \in \mathcal{Z}$ such that $g_1(z) = g_2(z) = t$, then JM is equivalent to SI.
Proof: see Appendix A5.

Corollary A1.1 *If for every $t \in [0,1]$ there exists $z \in \mathcal{Z}$ such that $P[D{=}n|Z{=}z] = t$ and $P[D{=}c|Z{=}z] = 0$, then OM and UM are equivalent to SI.*

Proposition A1 reveals that Assumption JM introduces a sequential threshold crossing structure on judge decisions: they first assign each individual a rank of not being convicted (U_1) and not being incarcerated (U_2), and then convict and additionally incarcerate convicted individuals with ranks below their thresholds. However, through (A23) and (A24), Proposition A1 reveals that Assumption JM also introduces an additional restriction on how judges allow individuals to differ in their rankings across the two decision margins. To better see the content of this restriction, Figure A3(a) graphically illustrates the inadmissible area of rankings in the case of a single judge who does not convict any defendants. Here we can see that while rankings such as (u_1', u_2') and (u_1'', u_2'') are permitted, those such as (u_1'', u_2') or (u_1', u_2'') that increase the rank of one decision margin relative to the other are not. This highlights that the restriction can imply that whenever a judge assigns an individual a high rank in one margin, they necessarily must do so in the other.

As illustrated in Figure A3(b), the restriction becomes stronger in the presence of more judges as the area of inadmissible rankings increases. Proposition A1 sharpens this observation when there is sufficient continuous variation in judges' thresholds. It shows that in this case, Assumption JM imposes a homogeneous rank for the incarceration and conviction margins. This is a strong restriction on judge behavior. For example, consider an individual plausibly guilty of certain petty misdemeanor crimes. For such an individual, judges may assign a high rank of being convicted but not necessarily incarcerated. A homogeneous rank for the two decision margins, however, rules out such realistic scenarios. Since JM is weaker than OM and UM, this also implies that these assumptions also converge to SI.

While there are no judges in our setting who assign no defendants to treatment c, the above proposition suggests that tests of the validity of SI may shed light on the reasonableness of models of ordered and unordered monotonicity. In Appendix A6, we show that under SI, for any characteristic X, the Wald estimands between two judges on the outcomes XD_d and treatments D_d for $d \in \{n, p\}$ are bounded. This follows because the characteristics X of treatment d individuals are exactly controlled by changes in the treatment share for these outcome moments in this model. We adopt a semiparametric test developed in Frandsen, Lefgren and Leslie (2023) designed for the case of judge comparisons and find that we reject this test for some covariates, indicating that the data appears to be inconsistent with SI.

A5 Proofs of index representation propositions

A5.1 Proof of Proposition A1

Following Vyltacil (2002), we have that Assumption JM can be equivalently written as

$$
\begin{aligned}
1\{D_{cp}(z)=1\} &= 1\{U_1 \leq g_1(z)\}\ , \\
1\{D_p(z)=1\} &= 1\{U_2 \leq g_2(z)\}\ ,
\end{aligned}
$$

where $U_1, U_2 \sim U[0,1]$, and $g_1(z) = 1 - P(D(z)=n)$ and $g_2(z) = P(D(z)=p)$. As

$$
D(z) = p1\{D_p(z)=1,\ D_{cp}(z)=1\} + c1\{D_p(z)=0,\ D_{cp}(z)=1\} + n1\{D_{cp}(z)=0\}\ ,
$$

the threshold crossing equation in (A22) then directly follows. Next, to obtain the restriction in (A23) and (A24), observe that since logically $P[D_{cp}(z)=0,\ D_p(z)=1]=0$, it follows that

$$
P[U_1 > g_1(z),\ U_2 \leq g_2(z)] = 0\ . \tag{A25}
$$

Moreover, since $P(D(z)=n)+P(D(z)=c)+P(D(z)=p)=1$ and $P[D(z)=c]=P[U_1 \leq g_1(z),\ U_2 > g_2(z)]$, we have

$$
P[U_1 \leq g_1(z),\ U_2 > g_2(z)] = g_1(z) - g_2(z) \tag{A26}
$$

This completes the proof.

A5.2 Proof of Corollary A1.1

Since $\mathcal{Z}$ is such that for every $t \in [0,1]$ there exists $z \in \mathcal{Z}$ such that $g_1(z) = g_2(z) = t$, it directly follows from (A23) and (A24) that

$$
\begin{aligned}
P[U_1 > t,\ U_2 \leq t] &= 0\ , \\
P[U_1 \leq t,\ U_2 > t] &= 0,
\end{aligned}
$$

for all $t \in [0,1]$. This implies $P(U_1 = U_2) = 1$.

A5.3 Proof of Proposition 2

The proof is identical to the first part of that of Proposition A1.

A6 Testable implications of model assumptions across judges

We show testable implications from the single-index (SI) and latent monotonicity (LM) assumptions on binary judge comparisons. Consider any defendant characteristic X, although we will assume it ranges between 0 and 1 for simplicity.[2] We use expressions for Wald estimands over XD_p instrumenting for D_p, which provides information on complier characteristics of incarcerated defendants, and over XD_n, instrumenting for D_n, providing information on dismissed defendants. By examining treatment-specific characteristics of defendants, we isolate treatment margins that are restricted by the underlying model assumptions and provide bounds for the estimands. We compare judges $Z = 1$ and $Z = 0$, and let $p_{ij} = P[D(1) = i, D(0) = j]$.

The Wald estimands can be rewritten as:

$$\begin{aligned}
\gamma_p &= \frac{E[XD_p|Z=1] - E[XD_p|Z=0]}{E[D_p|Z=1] - E[D_p|Z=0]} \\
&= \frac{E[X|D(1)=p, D(0)=n]p_{pn} + E[X|D(1)=p, D(0)=c]p_{pc}}{p_{pn} + p_{pc} - p_{np} - p_{cp}} \\
&+ \frac{E[-X|D(1)=n, D(0)=p]p_{np} + E[-X|D(1)=c, D(0)=p]p_{cp}}{p_{pn} + p_{pc} - p_{np} - p_{cp}}
\end{aligned}$$

and

$$\begin{aligned}
\gamma_n &= \frac{E[XD_n|Z=1] - E[XD_n|Z=0]}{E[D_n|Z=1] - E[D_n|Z=0]} \\
&= \frac{E[X|D(1)=n, D(0)=p]p_{np} + E[X|D(1)=n, D(0)=c]p_{nc}}{p_{np} + p_{nc} - p_{pn} - p_{cn}} \\
&+ \frac{E[-X|D(1)=p, D(0)=n]p_{pn} + E[-X|D(1)=c, D(0)=n]p_{cn}}{p_{np} + p_{nc} - p_{pn} - p_{cn}}
\end{aligned}$$

The single-index assumption implies

$$\begin{aligned}
0 \leq \gamma_p \leq 1 \\
0 \leq \gamma_n \leq 1.
\end{aligned}$$

To see the relation for γ_p, assume $P[D_p|Z = 1] \geq P[D_p|Z = 0]$. This implies $g_p(1) \geq g_p(0)$, which is the sole cutoff threshold for D_p, and consequently, $p_{np} = p_{cp} = 0$, i.e. no defiers move out of treatment p. Finally X is bounded between 0 and 1, so the numerator is bounded $[0, p_{pn}+p_{pc}]$, establishing the result. For the relation for γ_d, assume $P[D_n|Z = 1] \geq P[D_n|Z = 0]$. This implies $g_n(1) \geq g_n(0)$, and consequently $p_{pn} = p_{cn} = 0$.

The latent monotonicity assumption implies

$$\begin{aligned}
-\infty \leq \gamma_p \leq \infty \\
0 \leq \gamma_n \leq 1.
\end{aligned}$$

[2] If X has a wider range, the bounds on the Wald estimands we derive simply scale by the size of this range. In addition, we note that X could be endogenous to the treatment but is not required.

This model defines the same cutoffs g_c as in the single-index case, and consequently $p_{pn} = p_{cn} = 0$ for the D_n moments, leading to the relation on γ_n. For γ_p, there is no restriction on compliance types. To see that γ_p is unbounded, consider a condition in which the first stage denominator is 0, but the numerator is positive or negative. This is possible given that $p_{np} + p_{nc} - p_{pn} - p_{cn} \in [-1, 1]$ and the weighted average in the numerator is not restricted based on the first stage coefficients. Intuitively, multiple judge thresholds control entry into treatment p under LM, so knowing the share does not restrict the potential for two-way flows into and out of this treatment across judges.

The above conditions can be tested using the methods developed in Frandsen, Lefgren and Leslie (2023). This method was designed to test single-treatment IV assumptions for exclusion and monotonicity violations. It can instead be used to test the implications of the single-index or latent monotonicity assumptions on the appropriately defined outcome and treatment moments discussed here, as the between-judge slope conditions are similar.[3] A clear benefit of this approach is that the inference procedure is designed to account for estimation error in the judge propensities.

Table A10 presents results from the semi-parametric "fit" test across court-year cells. The χ^2 test statistics are aggregated across court-year cells to provide a joint test, since the test statistics and associated DOFs can be summed under the independence assumption. We run tests with D_n and D_p interacted with two covariates, any past charge and any future charge over the 5 years post-filing. Columns (1) and (2) show that we cannot reject the test on the D_n moments for either covariate ($p = 1$). This moment condition is the only test of the latent montonicity assumption, and hence the data provides some support for the underlying assumption.

We also test the D_p moments, which uniquely are implied by the single-index model. Column (3) shows we reject the test for the variable of any past charge, $\chi^2(DOF) = 1349(1208), p = 0.005$, while column (4) shows we cannot reject for any future charges $\chi^2(DOF) = 1204(1208), p = 0.53$. Rejecting the test on the D_n moments indicates that the data appear inconsistent with the implications of the single-index assumption. Together, these model implications and associated tests provide evidence against the single-index model and instead provide some support that the data are consistent with the latent monotonicity assumption.

[3] This test may be conservative as the Frandsen, Lefgren and Leslie (2023) test is designed to bound the between-judge differences to be -1 and 1 for an outcome with a 0 to 1 range.

A7 Identification in Humphries et al. (2023)

Another approach to identification of multiple treatment effects in examiner designs is Humphries et al. (2023), hereafter HOSSD.[4] In contrast to the standard examiner assignment design—and in contrast to our method—their method relies exclusively on covariates for identification of the first stage. In particular, they use these covariates to identify the coefficients on judge indicators in a multinomial choice model, and then use these coefficients as instruments in the second-stage model of Mountjoy (2022), hereafter M22.

In this section, we elaborate on the differences in our approaches, focusing on the first stage. We then implement their method in our data, as well as a partially identified version of their approach that differs only in not using covariates for identification. We find that the HOSSD method produces estimates that often lie *outside* the semiparametric bounds, suggesting the identification conditions are not satisfied (in our setting), which can lead to incorrect conclusions.

A7.1 Identification in HOSSD relies on additional separable regressors

HOSSD uses a multinomial choice model of judge decision-making, where

$$D(z) = \underset{d \in \{n,c,p\}}{\text{argmax}}\, U_{idzx} \tag{A27}$$

$$U_{idzx} = \begin{cases} 0 & \text{if} \;\; d = n, \\ W_i\beta_{dx} - g_{dx}(z) + \varepsilon_{idzx} & \text{if} \;\; d \in \{c, p\} \end{cases} \tag{A28}$$

and where the distribution of ε_{idzx} is known up to a finite-sized parameter vector of length α. β_{dx} represents the effect of potentially individual-level covariates W_i on decisions in court-year x, and $g_{dx}(z)$ represent the judge-court-year-specific thresholds. All models are separately estimated by court-year x.[5]

To understand how identification works in their setting, it is informative to simplify their model to a case with no covariates and a single court-year x, and where judges are unconditionally randomly assigned. The removal of covariates in similar settings is typically innocuous as they are uncorrelated to the instruments by design; classic expositions of instrumental variables (e.g. Imbens and Angrist, 1994) do not even include a discussion of covariates. We can write the distribution of U as

$$U_{idzx} = \begin{cases} 0 & \text{if} \;\; d = n, \\ -g_{dx}(z) + \varepsilon_{idzx} & \text{if} \;\; d \in \{c, p\} \end{cases} \tag{A29}$$

This choice equation gives rise to the following relationship between the judge thresholds

[4] This note refers to their July 2023 version available here.

[5] Their specification on page 33 does not include individual controls W_i, but we confirmed they are included in a discussion with the authors.

g and the judge-specific choice propensities:

$$\begin{aligned} P[D{=}n|Z{=}z, X{=}x] &= \int_{-\infty}^{g_{cx}(z)} \int_{-\infty}^{g_{px}(z)} f(u_1, u_2)\ du_2\ du_1 \\ P[D{=}c|Z{=}z, X{=}x] &= \int_{g_{cx}(z)}^{\infty} \int_{-\infty}^{g_{px}(z)-g_{cx}(z)+u_1} f(u_1, u_2)\ du_2\ du_1 \\ P[D{=}p|Z{=}z, X{=}x] &= \int_{g_{px}(z)}^{\infty} \int_{-\infty}^{g_{cx}(z)-g_{px}(z)+u_2} f(u_1, u_2)\ du_1\ du_2 \end{aligned} \tag{A30}$$

where F is the distribution of ε. However, from this representation, it is clear that g cannot be identified unless this distribution is known. Indeed, as in our model, each distribution F implies a different mapping between the observed choice propensities and thresholds g that rationalize them. This can be seen more concretely in Figure A2, which takes F as a normal distribution with unit variances and correlation ρ. For a judge that assigns 30, 20, and 50% of defendants to treatments n, c, and p, respectively, we plot $\big(g_{1x}(z), g_{2x}(z)\big)$ for each $\rho \in [0, 0.2, 0.4, 0.6, 0.8, 1]$. The judge thresholds g vary considerably with ρ, illustrating that in the standard no-covariates case, the HOSSD first stage cannot be identified without prior knowledge of F.

This poses a threat to the interpretation of any such model, because F is precisely the parameter that governs substitution behavior across judges as well as the existence and size of the compliance groups. One option—which we pursue in this paper—is to accept that the first stage is only partially identified and that, in turn, many of the parameters of interest are also only partially identified. HOSSD instead relies on the existence of an additional set of regressors to point-identify g under additional restrictive assumptions on the data-generating process.

These regressors are represented by W_i in (A28). Since they are separable from the judge thresholds g, they are implicitly assumed to shift each judge's threshold in index space by the same amount. Combined with a parametric assumption on the distribution of ε, this allows identification of all model parameters. For a given court-year, if there are $|Z|$ judges and the W's are discrete with $|W|$ support points, there are $2|Z||W|$ moments $P[D{=}d|Z{=}z, W{=}x]$ but only $2(|W|-1)+2|Z|+a$ parameters, where a represents the number of parameters that govern F. Without regressors, i.e. when $|W|{=}1$, there are more parameters than moments. Adding even one regressor, in principle, allows identification.

In this light, however, identification in HOSSD should be understood as arising *because* of the existence of W and the assumption of separability. This is in contrast to standard instrumental variables approaches, where W typically consists of stratification cell indicators motivated by the treatment assignment process and identification is driven by instrument differences within cells. The selection of W is therefore a key design choice in HOSSD. Unfortunately, however, theory provides no guidance on how to choose these variables. This makes it difficult for the researcher to understand whether any given W satisfies the separability condition, and to adjudicate between results from models that rely on different W's for identification.

A7.2 HOSSD estimates lie outside the semiparametric bounds

To understand the implications of HOSSD's separability assumption, we implement their method as faithfully as possible using our data. We then compare it to a version of their approach that does not use covariates for identification, and as a result is only partially identified. The covariate-identified estimates lie mostly outside the semiparametric bounds, suggesting that at least in our setting, the identification conditions required by HOSSD are not satisfied.

We mirror their first stage as described in (A28), estimating mixed logit models where ε consists of the sum of a standard logistic distribution and a normally-distributed random effect. We allow the random effects for c and p to be correlated and have unrestricted variances, as in their preferred specification. X includes number of previous charges, charge type (drugs, property, sex, violent, family, other), average sentence length of charges, sex, and age. As in their implementation, the models are estimated separately by court and 3-year groupings, delivering estimated judge-year coefficients for the middle year.

To estimate the model of outcomes, we apply code from M22 almost directly. We use the judge-year logit coefficients as instruments for conviction and incarceration. In all regressions, we control for court-year fixed effects and the same observable characteristics as in the first stage. The linear regressions are weighted by an Epanechnikov kernel with bandwidth of 3, and we take the point of evaluation to be the mean of the residualized instruments, all following HOSSD.

Table A11 presents the results. This method produces: (a) the effect of conviction c relative to not guilty n, $MTE_{n\to c}$, which they call the "labeling" effect, in columns (1)-(3) and (b) the effect of c relative to prison p, $MTE_{p\to c}$, which they call the "decarceration" effect, in columns (4)-(6). In the full sample, column (1) shows large reductions in future charges from the labeling effect ($\beta = -0.49$) and column (4) indicates a large increase in charges from the decarceration effect ($\beta = 0.38$). The effects are broadly similar results for defendants charged in felony or misdemeanor courts.

Taken at face value, the results from the HOSSD method indicate a highly efficacious criminal justice system, wherein each additional sanction wields substantial crime-reducing power. This is surprising and conflicts with our preferred estimates, replicated in columns (2) and (5) and labeled KNP 2024. In the full sample, the HOSSD method estimates of both the labeling and decarceration effects are outside of our preferred estimate's 95% CIs.[6] While the preferred estimates of this paper (the KNP 2024 columns) feature prominent heterogeneity in the incarceration and conviction effects across felony and misdemeanor court defendants, the HOSSD method results do not.

The two methods differ in both first stage identification and subsequent treatment effect identification and estimation. To isolate only the difference in first stage identification, in columns (3) and (6) we adopt the choice model and M22 approach used in HOSSD while

[6] Specifically, the overall effect of conviction relative to incarceration estimated here, 0.38, is outside the bounds of our 95% confidence interval $(0.143, 0.322)$. The estimated conviction effects -0.49 are also outside the 95% CIs $(-0.315, 0.217)$.

partially identifying the first stage as in this paper. Similar to their approach, we assume an unordered choice model with unobservables distributed according to a bivariate logistic distribution with unit variance marginals and correlation in each court-year x induced by an unknown $\theta_x \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, the association parameter in the AMH copula (Ali, Mikhail and Haq, 1978). For each value of $\theta = \{\theta_x\}_{x\in\mathcal{X}}$, we identify the judge thresholds from the data and then apply the M22 method to estimate treatment effects.[7] The bounds are given by the union over the possible values of θ.

Given that this approach imposes fewer assumptions, we expect the HOSSD estimates to be within the estimated bounds. Instead, the HOSSD estimates are outside the semiparametric bounds for 5 of the 10 subsamples. We can reject that the estimates are inside the bounds for 4 at the 95% level.[8] This indicates that the identification conditions, notably separability, in HOSSD are unlikely to be satisfied in our setting and may result in incorrect conclusions.

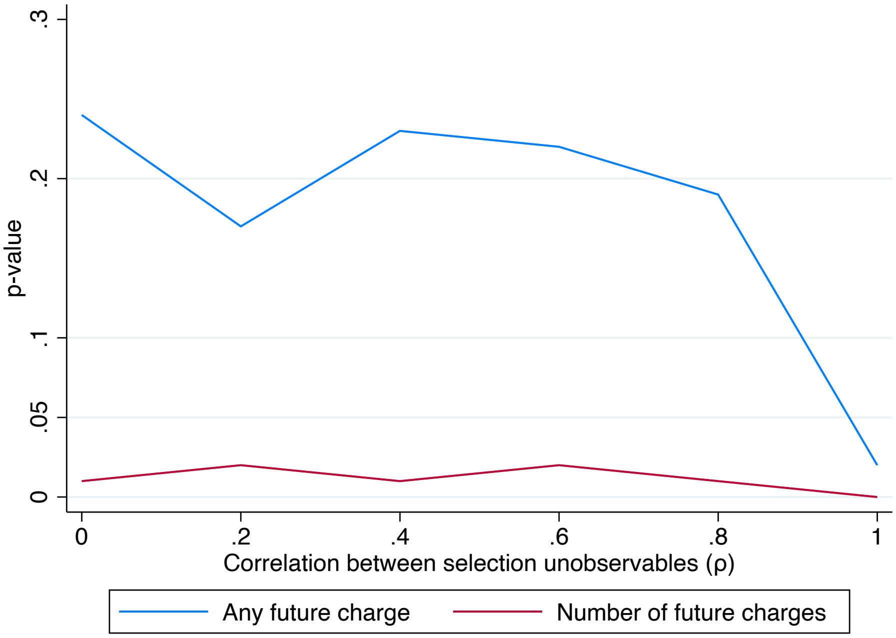
Aside from the issues with first stage identification highlighted above, there are several additional explanations why the results using the methodology from HOSSD may diverge from the baseline estimates in this paper. First, the "comparable compliers" assumption of M22 that is also required in HOSSD may not be satisfied. Tests of this assumption that are detailed in M22 require additional covariates not used for identification; however, as in HOSSD, we use nearly every covariate available for identification of the first stage model, with no obvious candidate to leave for testing. Second, M22 requires continuous variation in the instruments. Since judges are discrete, it is not clear that these local identification results can even be applied. Finally, our baseline estimates use 2SLS weights to aggregate effects across compliers, rather than report the effects for a particular marginal individual as in HOSSD.

[7] Since there are 114 court-years in our data and 6 possible values of θ_x in each court-year, directly estimating treatment effects in each of the 114^6 possible realizations of θ is not feasible. Instead, we note the coefficients from the linear regressions required for M22 are variance-weighted averages of the court-year-specific estimates. We estimate the weights and coefficients for each of the 6×114 θ_c-court-years, and then search over θ to minimize (maximize) the target parameters.

[8] This test is conservative. Denoting the semiparametric effect as a function of θ_x as $\beta(\cdot)$ and the corresponding HOSSD effect as β, in each bootstrap iteration b we calculate the lower bound on the difference as $\min_{\theta_x} \beta_b(\theta_x) - \beta_b$. We estimate the upper bound in the corresponding way, and use these estimated bounds across bootstrap samples to construct a 95% confidence interval. Four of ten confidence intervals do not include zero.

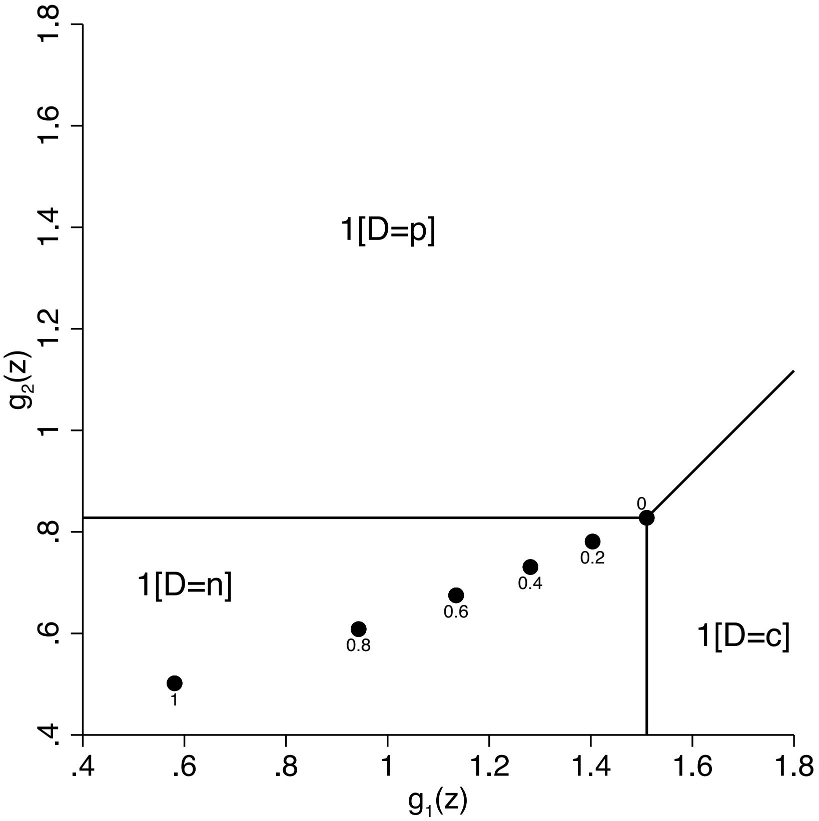
A8 Appendix Figures

Figure A1: p-values for null of matching 2SLS estimate of effect of incarceration, by ρ



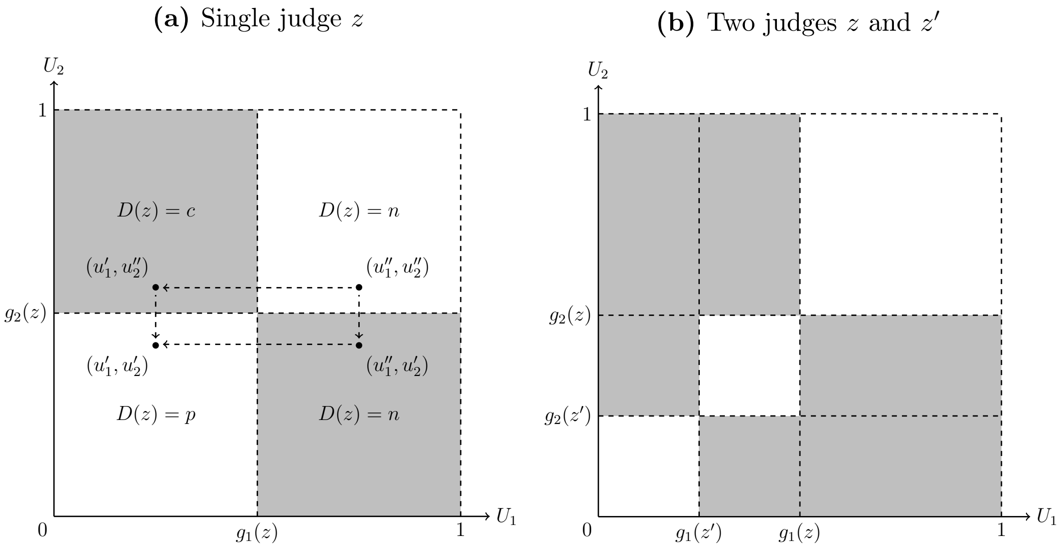
This figure shows the estimated p-values for a test of the null hypothesis that the structural model recovers the 2SLS coefficient from a regression of the given outcome on incarceration, where incarceration is instrumented by judge indicators. Distribution of 2SLS coefficients under null estimated via a boostrap with 200 draws.

Figure A2: Judge thresholds g in multinomial choice model when U is normally distributed with varying correlation ρ



For $\rho \in [0, 0.2, 0.4, 0.6, 0.8, 1]$, this graph plots $(g_1(z), g_2(z))$ for a judge that assigns 30, 20, and 50% of defendants to treatments n, c, and p, respectively.

Figure A3: Inadmissable regions space of unobservables under JM


(a) Single judge z

(b) Two judges z and z'

This figure displays different regions implied by Proposition A1 for an example with a single judge z with $g_1(z) = g_2(z)$ and how the inadmissible region (corresponding to the shaded gray areas) increases with an additional judge z' with thresholds equal to half of those of z. Note that we drop the conditioning on $x \in \mathcal{X}$ for convenience.

A9 Appendix Tables

Table A1: Effects of conviction and incarceration on future binary criminal justice outcomes

				Never prev. convicted	
	All	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)	(5)
			Panel A: Charged over next 5 years		
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.060, -0.047]	[-0.100, -0.083]	[0.005, 0.021]	[-0.084, -0.062]	[-0.006, 0.008]
	(-0.077, -0.032)	(-0.120, -0.064)	(-0.022, 0.037)	(-0.115, -0.038)	(-0.036, 0.038)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[0.143, 0.736]	[0.144, 0.704]	[0.143, 0.751]	[0.097, 1.096]	[0.230, 0.832]
	(0.034, 1.267)	(-0.586, 1.993)	(0.031, 1.278)	(-0.998, 3.190)	(0.079, 1.416)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[-0.033, 0.020]	[-0.052, 0.032]	[-0.024, 0.014]	[-0.037, 0.029]	[0.022, 0.045]
	(-0.065, 0.060)	(-0.090, 0.100)	(-0.074, 0.071)	(-0.084, 0.094)	(-0.028, 0.106)
			Panel B: Convicted over next 5 years		
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.066, -0.050]	[-0.096, -0.079]	[-0.018, 0.003]	[-0.083, -0.061]	[-0.030, -0.012]
	(-0.083, -0.035)	(-0.118, -0.059)	(-0.047, 0.028)	(-0.113, -0.035)	(-0.062, 0.013)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[0.160, 0.874]	[0.133, 0.768]	[0.182, 0.921]	[0.069, 0.698]	[0.237, 0.942]
	(0.067, 1.376)	(-0.525, 2.062)	(0.074, 1.422)	(-1.388, 2.784)	(0.108, 1.539)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[-0.005, 0.049]	[-0.051, 0.027]	[0.018, 0.060]	[-0.020, 0.037]	[0.057, 0.087]
	(-0.034, 0.088)	(-0.107, 0.094)	(-0.033, 0.117)	(-0.086, 0.103)	(0.009, 0.142)
Weight on $c \to p$ effect	[0.977, 1.000]	[0.991, 1.000]	[0.947, 1.000]	[0.989, 1.000]	[0.966, 1.000]
Weight on $n \to p$ effect	[0.000, 0.054]	[0.000, 0.037]	[0.000, 0.086]	[0.000, 0.043]	[0.000, 0.074]
Weight on $n \to c$ effect	[0.065, 0.088]	[0.030, 0.045]	[0.131, 0.169]	[0.042, 0.058]	[0.124, 0.152]

This table reports treatment effects of conviction and incarceration, aggregated using the weights from a 2SLS regression with incarceration as the treatment and judge dummies as the instruments. MTRs are approximated by a second-degree polynomial in u_1 and u_2 as specified in Section 5.5. Bounds in square brackets and 95% confidence intervals calculated using Bei (2023) in parentheses.

Table A2: Effects of conviction and incarceration on future criminal justice outcomes, by prior record

	No prev. felony or misdemeanor convictions		No prev. felony convictions	
	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)
Panel A: Number of charges over next 5 years				
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.330, -0.271]	[0.002, 0.026]	[-0.333, -0.265]	[-0.037, 0.024]
	(-0.442, -0.166)	(-0.117, 0.142)	(-0.426, -0.195)	(-0.122, 0.097)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[-2.770, -0.079]	[-0.324, 0.313]	[-4.237, -0.068]	[0.851, 3.336]
	(-10.449, 4.909)	(-2.800, 2.155)	(-11.125, 2.562)	(0.377, 5.959)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[-0.074, 0.104]	[0.173, 0.253]	[-0.187, 0.015]	[0.165, 0.351]
	(-0.314, 0.305)	(-0.023, 0.463)	(-0.417, 0.227)	(-0.016, 0.492)
Conviction effect ($\Delta^*_{n \to c}$)	[-0.125, -0.039]	[0.173, 0.712]	[-0.652, -0.187]	[0.165, 0.813]
	(-0.834, 0.756)	(-0.004, 1.116)	(-1.558, 0.304)	(-0.009, 1.223)
Panel B: Number of convictions over next 5 years				
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.279, -0.212]	[-0.050, -0.039]	[-0.272, -0.190]	[-0.074, -0.013]
	(-0.376, -0.110)	(-0.195, 0.099)	(-0.383, -0.088)	(-0.259, 0.164)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[-0.012, 0.354]	[0.328, 0.396]	[-0.949, 0.278]	[1.166, 3.870]
	(-7.898, 7.873)	(-2.731, 3.388)	(-9.108, 7.211)	(0.302, 7.333)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[-0.021, 0.132]	[0.275, 0.373]	[-0.141, 0.088]	[0.411, 0.625]
	(-0.267, 0.320)	(0.068, 0.606)	(-0.470, 0.313)	(0.189, 0.884)
Conviction effect ($\Delta^*_{n \to c}$)	[-0.297, -0.015]	[0.275, 0.882]	[-0.950, -0.043]	[0.411, 1.175]
	(-1.064, 0.445)	(0.089, 1.292)	(-1.725, 0.246)	(0.203, 1.636)
Weight on $c \to p$ effect	[0.992, 1.000]	[0.960, 1.000]	[0.989, 1.000]	[0.966, 1.000]
Weight on $n \to p$ effect	[0.000, 0.045]	[0.000, 0.078]	[0.000, 0.043]	[0.000, 0.074]
Weight on $n \to c$ effect	[0.050, 0.067]	[0.126, 0.158]	[0.042, 0.058]	[0.124, 0.152]
Weight on combined $n \to c$ effect	[0.067, 0.095]	[0.158, 0.205]	[0.058, 0.085]	[0.152, 0.198]

This table reports treatment effects of conviction and incarceration, aggregated using the weights from a 2SLS regression with as the treatment and judge dummies as the instruments. MTRs are approximated by a second-degree polynomial in u_1 and u_2 as specified in Section 5.5. Bounds in square brackets and 95% confidence intervals calculated using Bei (2023) in parentheses.

Table A3: Effects of conviction and incarceration on future criminal justice outcomes, conviction 2SLS-weighted

				Never prev. convicted	
	All	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)	(5)
Panel A: Number of charges over next 5 years					
Incarceration rel. to conviction ($\Delta_{c\to p}$)	[0.040, 0.252]	[0.069, 0.131]	[0.017, 0.318]	[-0.143, 0.037]	[-0.013, 0.224]
	(-0.076, 0.347)	(-0.145, 0.371)	(-0.142, 0.435)	(-0.329, 0.317)	(-0.201, 0.357)
Incarceration rel. to not guilty ($\Delta_{n\to p}$)	[0.486, 2.897]	[0.183, 0.541]	[0.591, 4.012]	[-4.315, -0.076]	[1.002, 4.570]
	(0.158, 5.279)	(-4.455, 4.821)	(-0.004, 6.848)	(-9.261, 0.495)	(0.575, 7.600)
Conviction rel. to not guilty ($\Delta_{n\to c}$)	[-0.189, 0.017]	[-0.223, 0.026]	[-0.194, 0.014]	[-0.103, 0.058]	[0.026, 0.131]
	(-0.285, 0.126)	(-0.505, 0.219)	(-0.365, 0.200)	(-0.315, 0.216)	(-0.093, 0.302)
Panel B: Number of convictions over next 5 years					
Incarceration rel. to conviction ($\Delta_{c\to p}$)	[-0.028, 0.193]	[0.145, 0.344]	[-0.130, 0.210]	[0.060, 0.304]	[-0.111, 0.198]
	(-0.198, 0.304)	(-0.008, 0.669)	(-0.352, 0.377)	(-0.123, 0.533)	(-0.368, 0.369)
Incarceration rel. to not guilty ($\Delta_{n\to p}$)	[0.698, 4.048]	[-0.650, 0.382]	[0.962, 5.978]	[-3.607, -0.009]	[1.428, 6.335]
	(0.131, 7.406)	(-6.437, 5.136)	(0.053, 10.364)	(-9.614, 2.400)	(0.373, 10.820)
Conviction rel. to not guilty ($\Delta_{n\to c}$)	[0.054, 0.254]	[-0.248, 0.053]	[0.104, 0.328]	[-0.033, 0.140]	[0.246, 0.342]
	(-0.084, 0.411)	(-0.563, 0.275)	(-0.076, 0.550)	(-0.274, 0.318)	(0.071, 0.556)
Weight on $c \to p$ effect	[0.334, 0.444]	[0.316, 0.438]	[0.341, 0.446]	[0.279, 0.358]	[0.268, 0.339]
Weight on $n \to p$ effect	[0.000, 0.237]	[0.000, 0.366]	[0.000, 0.190]	[0.000, 0.292]	[0.000, 0.160]
Weight on $n \to c$ effect	[0.768, 1.005]	[0.639, 1.005]	[0.816, 1.006]	[0.714, 1.006]	[0.845, 1.006]

This table reports treatment effects of conviction and incarceration, aggregated using the weights from a 2SLS regression with conviction as the treatment and judge dummies as the instruments. MTRs are approximated by a second-degree polynomial in u_1 and u_2 as specified in Section 5.5. Bounds in square brackets and 95% confidence intervals calculated using Bei (2023) in parentheses.

Table A4: Effects of conviction and incarceration on future criminal justice outcomes measured after 7 years

				Never prev. convicted	
	All	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)	(5)
Panel A: Number of charges over next 7 years					
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.214, -0.174]	[-0.350, -0.316]	[0.010, 0.091]	[-0.298, -0.206]	[-0.015, 0.033]
	(-0.287, -0.107)	(-0.446, -0.238)	(-0.126, 0.206)	(-0.424, -0.093)	(-0.152, 0.154)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[0.165, 2.171]	[-3.113, -0.467]	[0.725, 4.546]	[-2.395, 0.297]	[0.988, 3.532]
	(-0.544, 4.700)	(-9.831, 3.605)	(0.233, 6.873)	(-10.378, 5.588)	(0.337, 6.549)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[-0.176, 0.069]	[-0.376, -0.131]	[-0.070, 0.176]	[-0.227, 0.062]	[0.243, 0.414]
	(-0.336, 0.259)	(-0.779, 0.147)	(-0.277, 0.404)	(-0.440, 0.313)	(0.027, 0.660)
Panel B: Number of convictions over next 7 years					
Incarceration rel. to conviction ($\Delta_{c \to p}$)	[-0.295, -0.250]	[-0.405, -0.363]	[-0.132, -0.036]	[-0.303, -0.177]	[-0.039, 0.037]
	(-0.397, -0.145)	(-0.512, -0.264)	(-0.346, 0.172)	(-0.448, -0.047)	(-0.274, 0.260)
Incarceration rel. to not guilty ($\Delta_{n \to p}$)	[0.176, 2.805]	[-4.785, -0.751]	[0.946, 6.217]	[0.260, 1.363]	[1.251, 4.666]
	(-0.908, 6.340)	(-13.736, 4.167)	(0.114, 10.293)	(-9.055, 9.576)	(0.277, 8.485)
Conviction rel. to not guilty ($\Delta_{n \to c}$)	[0.085, 0.382]	[-0.468, -0.022]	[0.195, 0.598]	[-0.119, 0.164]	[0.531, 0.781]
	(-0.112, 0.621)	(-0.989, 0.321)	(-0.070, 0.976)	(-0.562, 0.434)	(0.227, 1.143)
Weight on $c \to p$ effect	[0.977, 1.000]	[0.990, 1.000]	[0.949, 1.000]	[0.989, 1.000]	[0.967, 1.000]
Weight on $n \to p$ effect	[0.000, 0.054]	[0.000, 0.038]	[0.000, 0.085]	[0.000, 0.044]	[0.000, 0.073]
Weight on $n \to c$ effect	[0.062, 0.084]	[0.030, 0.045]	[0.122, 0.158]	[0.040, 0.057]	[0.117, 0.144]

This table reports treatment effects of conviction and incarceration, aggregated using the weights from a 2SLS regression with incarceration as the treatment and judge dummies as the instruments. MTRs are approximated by a second-degree polynomial in u_1 and u_2 as specified in Section 5.5. Bounds in square brackets and 95% confidence intervals calculated using Bei (2023) in parentheses.

Table A5: Policy effects after 7 years

	Change in treatment shares			Effects on outcomes	
	n	c	p	N. charges	N. conv.
Panel A: Felony defendants					
Incarceration leniency ($g_2 \downarrow$)	0.000	[0.009, 0.010]	[-0.010, -0.009]	[0.329, 0.757]	[0.298, 0.834]
				(0.212, 1.061)	(0.191, 1.205)
Conviction leniency ($g_1 \downarrow$)	0.010	[-0.010, -0.007]	[-0.003, 0.000]	[-0.093, 0.219]	[-0.043, 0.155]
				(-0.362, 0.421)	(-0.376, 0.465)
No incarceration ($g_2 = 0$)	0.000	0.292	-0.292	[0.077, 0.308]	[0.127, 0.321]
				(-0.036, 0.377)	(0.006, 0.482)
No conviction ($g_1 = 0$)	0.874	-0.582	-0.292	[-11.842, -11.779]	[-4.454, -4.370]
				(-22.263, -1.352)	(-18.724, 9.902)
Homogenize judges $\big(g(z,x) = g(x)\big)$	0.000	0.000	0.000	[-0.224, 0.187]	[-0.216, 0.209]
				(-0.333, 0.409)	(-0.350, 0.482)
Panel B: Misdemeanor defendants					
Incarceration leniency ($g_2 \downarrow$)	0.000	[0.005, 0.010]	[-0.010, -0.005]	[-0.246, 0.205]	[-0.283, 0.368]
				(-0.386, 0.396)	(-0.514, 0.654)
Conviction leniency ($g_1 \downarrow$)	0.010	[-0.010, -0.008]	[-0.002, 0.000]	[-0.151, 0.185]	[-0.481, -0.003]
				(-0.362, 0.371)	(-0.791, 0.226)
No incarceration ($g_2 = 0$)	0.000	0.107	-0.107	[-0.035, 0.039]	[-0.051, 0.049]
				(-0.055, 0.070)	(-0.080, 0.094)
No conviction ($g_1 = 0$)	0.539	-0.433	-0.107	[-0.775, -0.766]	[-1.259, -1.245]
				(-1.500, -0.041)	(-2.108, -0.398)
Homogenize judges $\big(g(z,x) = g(x)\big)$	0.000	0.000	0.000	[-0.014, 0.283]	[0.008, 0.477]
				(-0.154, 0.386)	(-0.208, 0.629)
Panel C: Never-convicted misdemeanor defendants					
Incarceration leniency ($g_2 \downarrow$)	0.000	[0.005, 0.010]	[-0.010, -0.005]	[-0.179, 0.144]	[-0.279, 0.210]
				(-0.340, 0.353)	(-0.535, 0.550)
Conviction leniency ($g_1 \downarrow$)	0.010	[-0.010, -0.008]	[-0.002, 0.000]	[-0.295, -0.061]	[-0.598, -0.297]
				(-0.468, 0.113)	(-0.876, -0.062)
No incarceration ($g_2 = 0$)	0.000	0.089	-0.090	[-0.019, 0.015]	[-0.027, 0.023]
				(-0.035, 0.039)	(-0.052, 0.059)
No conviction ($g_1 = 0$)	0.536	-0.446	-0.090	[-0.704, -0.693]	[-1.189, -1.173]
				(-1.384, -0.013)	(-1.942, -0.420)
Homogenize judges $\big(g(z,x) = g(x)\big)$	0.000	0.000	0.000	[-0.011, 0.206]	[0.020, 0.346]
				(-0.182, 0.301)	(-0.210, 0.483)

Table reports the effects of a number of policy changes on recidivism. We analyze marginal changes, which shift judges' thresholds g by 0.01, as well as global changes. The change in treatment shares is the change from the given policy. The change in outcomes is rescaled by the number of defendants whose treatment is affected by the policy change for the marginal changes to assist in readability. Bounds are in square brackets and the outer edges of 95% confidence intervals in parentheses are calculated using Bei (2023).

Table A6: Effects of conviction and incarceration on future criminal justice outcomes, Ali-Mikhail-Haq unobservables

				Never prev. convicted	
	All	Fel.	Misd.	Fel.	Misd.
	(1)	(2)	(3)	(4)	(5)
			Panel A: Number of charges over next 5 years		
Incarceration rel. to conviction ($\Delta_{c\to p}$)	[-0.250, -0.232]	[-0.376, -0.370]	[-0.007, 0.053]	[-0.333, -0.310]	[-0.032, 0.023]
	(-0.307, -0.170)	(-0.418, -0.327)	(-0.139, 0.180)	(-0.383, -0.249)	(-0.151, 0.139)
Incarceration rel. to not guilty ($\Delta_{n\to p}$)	[0.220, 1.104]	[-0.301, 0.152]	[0.563, 2.126]	[-0.528, -0.069]	[0.846, 1.909]
	(-0.214, 1.899)	(-1.052, 1.308)	(0.084, 3.366)	(-1.555, 0.112)	(0.495, 3.204)
Conviction rel. to not guilty ($\Delta_{n\to c}$)	[-0.035, 0.066]	[-0.270, -0.104]	[0.068, 0.146]	[-0.147, -0.036]	[0.293, 0.343]
	(-0.216, 0.226)	(-0.636, 0.169)	(-0.131, 0.341)	(-0.458, 0.200)	(0.175, 0.457)
			Panel B: Number of convictions over next 5 years		
Incarceration rel. to conviction ($\Delta_{c\to p}$)	[-0.284, -0.261]	[-0.362, -0.351]	[-0.135, -0.067]	[-0.272, -0.251]	[-0.078, -0.013]
	(-0.358, -0.180)	(-0.370, -0.343)	(-0.290, 0.075)	(-0.375, -0.150)	(-0.224, 0.120)
Incarceration rel. to not guilty ($\Delta_{n\to p}$)	[0.331, 1.364]	[-0.395, 0.192]	[0.786, 2.623]	[0.039, 0.278]	[1.166, 2.398]
	(-0.267, 2.362)	(-1.459, 1.517)	(0.188, 4.381)	(-0.925, 1.736)	(0.745, 4.130)
Conviction rel. to not guilty ($\Delta_{n\to c}$)	[0.195, 0.345]	[-0.379, -0.101]	[0.446, 0.556]	[-0.141, 0.016]	[0.540, 0.620]
	(0.000, 0.515)	(-0.760, 0.181)	(0.242, 0.758)	(-0.456, 0.255)	(0.310, 0.838)
Weight on $c \to p$ effect	[0.977, 0.977]	[0.989, 0.992]	[0.947, 0.971]	[0.989, 0.991]	[0.966, 0.978]
Weight on $n \to p$ effect	[0.051, 0.054]	[0.023, 0.037]	[0.041, 0.086]	[0.023, 0.043]	[0.035, 0.074]
Weight on $n \to c$ effect	[0.065, 0.067]	[0.030, 0.039]	[0.131, 0.156]	[0.042, 0.051]	[0.124, 0.142]

This table reports treatment effects of conviction and incarceration, aggregated using the weights from a 2SLS regression with incarceration as the treatment and judge dummies as the instruments. MTRs are approximated by a second-degree polynomial in u_1 and u_2 as specified in Section 5.5. Bounds in square brackets and 95% confidence intervals calculated using Bei (2023) in parentheses.

Table A7: Response types $(D(z), D(z'))$ between a pair of judges $z, z' \in \mathcal{Z}$ with $D_{cp}(z) \geq D_{cp}(z')$

Assumption	Ordering	$(D(z), D(z'))$								
		(n,n)	(n,c)	(n,p)	(c,n)	(c,c)	(c,p)	(p,n)	(p,c)	(p,p)
Ordered	$D_p(z) \geq D_p(z')$	✓			✓	✓		✓	✓	✓
Unordered	$D_c(z) \geq D_c(z'),\ D_p(z) \geq D_p(z')$	✓			✓	✓		✓		✓
	$D_c(z) \geq D_c(z'),\ D_p(z) \leq D_p(z')$	✓			✓	✓	✓			✓
	$D_c(z) \leq D_c(z'),\ D_p(z) \geq D_p(z')$	✓				✓		✓	✓	✓
Single Index	$g_2(z) < g_1(z')$	✓			✓			✓	✓	✓
	$g_1(z') < g_2(z) < g_2(z')$	✓			✓	✓			✓	✓
	$g_2(z') < g_2(z)$	✓			✓	✓	✓			✓
Latent	$D_p^*(z) \geq D_p^*(z')$	✓			✓	✓		✓	✓	✓
	$D_p^*(z) \leq D_p^*(z')$	✓			✓	✓	✓	✓		✓

Treatments are n (not guilty or dismissed), c (convicted but not incarcerated), and p (incarcerated). Ordered refers to Assumption OM, and Unordered refers to Assumption UM, Single Index refers to Assumption SI and Latent refers to Assumption LM. Ordering refers to the additional monotonicity conditions imposed by the assumptions in addition to $D_{cp}(z) \geq D_{cp}(z')$, which is imposed by all four assumptions (in SI this implies that $g_1(z) \leq g_1(z')$). Note that the ordering $D_c(z) \leq D_c(z'),\ D_p(z) \geq D_p(z)$ doesn't logically exist when $D_{cp}(z) \geq D_{cp}(z)$.

Table A8: Treatment assignment by judge for each compliance group in single-index example

	Judge		
Group ℓ	0	1	2
$(0.25, 0.30]$	c	p	p
$(0.30, 0.35]$	n	p	p
$(0.35, 0.40]$	n	c	p
$(0.40, 0.74]$	n	c	c
$(0.75, 0.85]$	n	n	c

This table reports the treatment assignment under judge j for each of the treatment compliance types induced by the single-index treatment model defined in SI where the judges judges have incarceration and conviction thresholds of $(g_2(z), g_1(z)) \in \{(0.25, 0.30), (0.35, 0.75), (0.40, 0.85)\}$. This generates five compliance groups, which are listed along the rows.

Table A9: Weights on compliance groups in two-treatment 2SLS

	Wt. for ℓ's $t\to s$ effect in β_c coef. $(\phi_{st}^{\ell c})$						Wt. for ℓ's $t\to s$ effect in β_p coef. $(\phi_{st}^{\ell p})$					
Group ℓ	$n\to c$	$n\to p$	$c\to p$	$c\to n$	$p\to n$	$p\to c$	$n\to c$	$n\to p$	$c\to p$	$c\to n$	$p\to n$	$p\to c$
$(0.25, 0.30]$	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
$(0.30, 0.35]$	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
$(0.35, 0.40]$	0.2	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.4	0.2	0.0	0.0
$(0.40, 0.75]$	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0
$(0.75, 0.85]$	0.0	0.0	0.0	0.8	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0

This table reports the weights for the different compliance groups generated in a regression where both incarceration and conviction are instrumented for with judge assignment. We assume that treatment is determined by the single-index model defined in SI, and that there are three equally-likely judges who have incarceration and conviction thresholds of $(g_2(z), g_1(z)) \in \{(0.25, 0.30), (0.35, 0.75), (0.40, 0.85)\}$. This generates five compliance groups, with $u \in \{(0.25, 0.30], (0.30, 0.35], (0.35, 0, 40], (0.40, 0.75], (0.75, 0.85]\}$. $\phi_{st}^{\ell c}$ is the weight of the $t\to s$ effect for group ℓ in the coefficient on treatment c; $\phi_{st}^{\ell p}$ is the analogous weight for the p coefficient.

Table A10: Using the Frandsen-Lefgren-Leslie test to test model assumptions

	Test for single-index and latent monotonicity		Test for single-index	
	(1)	(2)	(3)	(4)
	Any past charge $\times D_n$	Any charges $\times D_n$	Any past charge $\times D_p$	Any charges $\times D_p$
χ^2	828	850	1337	1201
Deg. of freeom	1208	1208	1208	1208
p-value	1	1	.00532	.55
Observations	638,684	638,684	638,684	638,684

This table displays results from the semi-parametric Frandsen test to adjudicate between choice models. Outcomes of the form XD_n apply the Frandsen-Lefgren-Leslie test with judges instrumenting for D_n, and XD_p does so for D_p. The table reports results from the fit component of the Frandsen-Lefgren-Leslie test, applied with 3 knots and done separately across court-year cells. The chi-square test statistics and degrees of freedom are aggregated across cells to test the joint condition. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Comparison of different methods for estimating multiple treatment effects

	Labeling effect ($MTE_{n \to c}$)			Decarceration effect ($MTE_{p \to c}$)		
	HOSSD 2023	KNP 2024		HOSSD 2023	KNP 2024	
Choice model:	Util. max.	Latent mono.	Util. max.	Util. max.	Latent mono.	Util. max.
Identification:	X separability	Z	Z	X separability	Z	Z
Estimation:	Mountjoy 2022	KNP 2024	Mountjoy 2022	Mountjoy 2022	KNP 2024	Mountjoy 2022
	(1)	(2)	(3)	(4)	(5)	(6)
All	-0.494	[-0.185, 0.071]	[-0.896, -0.437]	0.381	[0.202, 0.246]	[0.799, 1.313]
	(-0.925, -0.063)	(-0.315, 0.217)	(-1.267, -0.156)	(0.254, 0.508)	(0.143, 0.322)	(0.616, 1.582)
Fel.	-0.545	[-0.304, -0.087]	[-0.817, -0.519]	0.509	[0.345, 0.379]	[0.762, 1.040]
	(-1.181, 0.091)	(-0.617, 0.167)	(-1.301, -0.083)	(0.396, 0.622)	(0.269, 0.461)	(0.565, 1.286)
Mun.	-0.520	[-0.124, 0.151]	[-0.940, -0.508]	0.289	[-0.071, 0.021]	[1.164, 2.755]
	(-0.972, -0.069)	(-0.305, 0.365)	(-1.333, -0.247)	(0.055, 0.523)	(-0.159, 0.147)	(0.836, 3.541)
Fel. (No prev. conv.)	-0.206	[-0.187, 0.015]	[-0.509, -0.125]	0.453	[0.265, 0.333]	[0.413, 0.754]
	(-0.893, 0.480)	(-0.417, 0.227)	(-1.194, 0.503)	(0.296, 0.610)	(0.195, 0.426)	(0.061, 1.159)
Mun. (No prev. conv.)	-0.290	[0.165, 0.351]	[-1.014, -0.371]	0.324	[-0.024, 0.037]	[1.121, 3.096]
	(-0.735, 0.154)	(-0.016, 0.492)	(-1.592, -0.079)	(0.073, 0.575)	(-0.097, 0.122)	(0.758, 4.396)

This table reports treatment effects of conviction relative to incarceration and to no punishment on the number of charges over the next 5 years. The KNP results are identical to the baseline 2SLS-weighted structural estimates reported in Table 4, while the HOSSD results use the utility maximization model of Humphries et al. (2023). Their baseline model, which relies on separability of covariates X from the instruments for identification, is estimated on our data and shown in columns (1) and (4). We also estimate the same choice model without relying on separability for identification. These partially identified results are shown in columns (3) and (6). Bounds in square brackets and 95% confidence intervals in parentheses.