Crime and Monetary Punishment*

Felipe Diaz Klaassen†

April 14, 2023

Most recent version

Abstract

This paper examines the causal effect of fines and court fees on criminal reoffending for low-level misdemeanants by leveraging the quasi-random assignment of judges in North Carolina. I find strong deterrent effects, with the imposition of any fines or fees reducing the likelihood of reoffending by 9 percentage points within two years of the original offense. Treatment effects are weakly negative for almost all defendant types, contradicting the hypothesis that financial obligations induce poorer defendants to engage in crimes entailing economic gains, although I do find evidence of increased financial distress for poorer defendants. Reductions in recidivism are driven almost entirely by defendants living in wealthier neighborhoods and first time defendants. Additionally, using generic machine learning methods to fully characterize treatment effect heterogeneity, I compare alternative allocations of fines and court fees and find significant improvements, reducing reoffending by 20%, when prioritizing fines and fees to wealthier defendants.

Keywords: Monetary sanctions, Legal financial obligations, LFO, Specific deterrence, Recidivism, Judge leniency, Inequality.

JEL Codes: H72, K14, K42.

*I thank my advisors Rebecca Thornton, Alex Bartik, Ben Marx, and Andrew Garin for their continual guidance and support. I also thank Jorge Lemus, Scott Cunningham, Julian Reif, David Albouy, Mark Borgshulte, and Anna Kyriazis for their thoughtful comments and discussions.

†Department of Economics, University of Illinois at Urbana-Champaign, diazkla2@illinois.edu

1 Introduction

Monetary sanctions are one of the most ubiquitous punishment alternatives within the United States' criminal justice system.[1] Every year, over 16 million people go to court in the United States and although relatively few defendants end up spending time in jail or prison, most end up owing money in criminal debt due to court fees, fines, and restitution.[2,3] Further, the growth of the criminal justice system in the past few decades and the accompanying rising operational costs have prompted state and local governments to turn increasingly towards imposing monetary sanctions as a source of additional revenue (Makowsky, 2019; Maciag, 2020).[4,5]

Concern with the impact these monetary sanctions may have on poorer defendants, with claims that fines and fees may actually incentivize individuals to commit crime (Harris et al., 2010; Harris, 2016; Natapoff, 2018), has prompted a number of states and localities to either restrict (Sledge, 2018; NBC Montana, 2018; Sibilla, 2022) or outright eliminate monetary sanctions altogether (Thadani, 2018).[6,7] At the same time, there's also been growing bipartisan support for the graduating of economic sanctions, taking into account a defendant's ability to pay (Colgan, 2017). Nevertheless, these monetary sanctions are meant to fulfill other practical functions within the criminal justice system besides revenue generation, such as deterrence and retribution, and since they are much less costly than imprisonment, understanding their impact—and the role they may play as an intermediate punishment between incarceration and probation—is of high relevance.[8]

This paper presents some of the first evidence of the causal effect of fines and court

[1]Monetary sanctions or legal financial obligations (LFOs), comprise fines, court fees, restitution and forfeitures. In my data, only fines and court fees are present, so I refer to monetary sanctions either as such, as LFOs, or as fines and fees throughout.

[2]Between 3 and 4 million felony cases and over 13 million misdemeanor cases, are filed annually in state and federal courts (Natapoff, 2018; Stevenson and Mayson, 2018).

[3]66% of prison inmates in the United States had court-imposed fines, fees, or restitution in 2004 (Bureau of Justice Statistics, 2004). Alabama, Arizona, California, Florida, Georgia, Illinois, Louisiana, Michigan, Missouri, New York, North Carolina, Ohio, Pennsylvania, Texas, and Virginia all impose fees upon conviction (Bannon et al., 2010).

[4]Between 1993 and 2012, total real annual criminal justice expenditures grew by 74 percent from $157 to $273 billion, with local spending comprising approximately half of total expenditures (Council of Economic Advisers, 2015).

[5]From 1991 to 2004, the fraction of inmates with any monetary sanction rose from 25 to 66 percent (Bureau of Justice Statistics, 2004).

[6]This is a possibility long recognized, even by the Supreme Court: "[T]he perverse effect of inducing the [debtor] to use illegal means to acquire funds...." (Bearden v. Georgia, 461 U.S. 660, 1983).

[7]In a University of Alabama survey Cook (2014) reports that 17% of surveyed probationers admitted to criminal activity for the purpose of paying economic sanctions.

[8]As fines carry mainly just collection costs, while imprisonment entails foregone earnings, restrictions on consumption and freedom, expenditures on guards, supervisory personnel, buildings, food, etc.

fees. I find that monetary sanctions reduce reoffending by 9 percentage points within two years of the original offense. I explore some potential mechanisms and test the hypothesis of whether financial obligations induce poorer defendants to engage in crimes entailing economic gains: I find that treatment effects seem to be driven by the updating of expectations of monetary punishment and that fines and court fees do not increase reoffending, even among the most poor. However, reductions in recidivism are driven almost entirely by defendants living in wealthier neighborhoods; furthermore, I do find some suggestive evidence of increased financial distress for poorer defendants. Additionally, using generic machine learning methods to fully characterize treatment effect heterogeneity, I compare alternative allocations of fines and court fees and find significant improvements, reducing reoffending by 20%, when prioritizing fines and fees to wealthier defendants.

Identifying the causal effects of monetary sanctions alone is difficult for three main reasons: first, high-quality data on total fees and fines assessed to defendants is difficult to obtain; second, in current practice, monetary sanctions are usually bundled with other common penalties such as community service, probation, parole, or incarceration (Harris et al., 2011); and third, the levying of fines and fees is likely correlated with unobservable criminality. To address the data needs I use administrative data from North Carolina between 2014 and 2019, that includes detailed information regarding the universe of criminal offenses, fines, and court fees. I focus on misdemeanor defendants sentenced in North Carolina's District Courts. I begin with administrative court case records covering all non-traffic low-level misdemeanor cases from January, 2013, to December, 2019. I then link defendants to criminal cases—misdemeanors and felonies—across all North Carolina Courts during the same seven year period, in order to track defendants' reoffending (thus, proxying criminal involvement with arrests).[9]

To avoid conflating fines and fees with other penalties, I focus my analysis on defendants who are, and can only be, levied fines and fees: in North Carolina criminal sanctions for lower-level defendants (without much prior involvement with the criminal justice system) are limited to fines and fees.[10]

Finally, to recover the causal effects I need exogenous variation in the probability of receiving a monetary sanction. I exploit the fact that in North Carolina's lower courts defendants are randomly assigned to judges that differ in their propensity to levy fines and fees, and recover a local average treatment effect (LATE) for individuals at the margin of

[9] Police officers are the charging agency in North Carolina so court records capture close to the universe of arrests (Rose, 2018).

[10] Current Operations and Capital Improvements Appropriations Act of 2013, §18B.13, 2013 N.C. Sess. Laws 995, 1303–04).

being levied a fine or court fee.[11]

I begin my analysis by measuring judges' propensity for levying fines or fees using a leave-out, residualized measure, based on all other cases that a judge has ruled on that do not involve a given defendant. This residualized measure considers leave-out measures of all covariates' estimates, thus correcting for the bias of propensity measure when the number of covariates is large.[12] I find significant variation in judge leniency, ranging from -0.925 to 1.070, with a standard deviation of 0.124. This measure is highly predictive of judge decisions regarding fines or fees, but uncorrelated with case and defendant characteristics. I find that going from the most lenient to the least lenient judge increases a defendant's likelihood of being levied a fine or a fee by 46 percentage points.

To assess the relevance and validity of the instrument, I perform a number of checks, all of which suggest that the instrument is strong, good as randomly assigned, and satisfies both the exclusion restriction and monotonicity. I also explore additional estimation strategies—such as 2SLS using all judge dummies, limited information maximum likelihood (LIML), JIVE, and Belloni et al. (2014)'s IV lasso—and the results remain qualitatively similar.[13]

I then use this leniency measure to instrument for whether defendants are levied fines or fees. I find that defendants receiving a fine or a fee are 9.1 percentage points less likely to reoffend, within two years, than defendants shown leniency. This is a 25 percent decrease with respect to a sample mean of 36.5 percent, and a 27 percent decrease with respect to the mean for "complier" defendants ($p < 0.05$). The results show an immediate drop in the likelihood of a new offense within the first three months after the disposition of the original offense. This effect grows larger through the first 9 months and then remains steady throughout. The decreased likelihood of reoffending comes from reductions in both felonies and misdemeanors, although estimates are much noisier for the latter.

To check whether defendants are incentivized to commit crimes in order to pay for these fines, I distinguish between financially and non-financially motivated reoffending and proxy for defendants' (unobserved) financial resources using their neighborhood income. Financially motivated offenses are those that entail economic gain, such as property crimes, robberies, or selling drugs.[14] Interestingly, I find that the deterrent effect of fines and fees is

[11]Other examples of papers exploiting this "judge fixed effects" research design include Kling, 2006, Mueller-Smith, 2015 and Bhuller et al., 2020, studying the effects of incarceration; Norris et al., 2021 estimating the impact of parental incarceration; Dobbie et al., 2018 describing the impact of pretrial detention; and Agan et al., 2021 looking at the effects of misdemeanor prosecution.

[12]This is the UJIVE estimator of (Kolesár, 2013).

[13]See Appendix C.4.

[14]I follow Tuttle (2019) and consider crimes to be financially motivated if they are property crimes (excluding those that do not entail economic gains, such as arson), drug offenses that involve their sale, manufacture, or distribution, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a

entirely for financially motivated offenses (10.3 p.p, $p < 0.01$). I find no effects of receiving a fine or a fee on recidivism involving non-financially motivated offenses; the point estimate is small and statistically insignificant (1.1 p.p, SE=0.040). Estimates are qualitatively similar if considering property and non-property offenses: coefficients maintain the same sign and are of similar magnitudes and significance. These effects are driven entirely by defendants living in wealthier neighborhoods. Defendants living in census tracts in the third and fourth quartiles of median household income are 13.7 and 12.1 percentage points less likely to engage in financially motivated crime 2 years following their disposition. This corresponds to 169 ($p < 0.05$) and 161 ($p < 0.1$) percent decreases with respect to their respective sample means. I find no effects for defendants living in the first and second quartiles. I also split the sample according to the federal poverty line in North Carolina.[15] Fines and court fees reduce the likelihood of reoffending by 9.7 percentage points for defendants living in census tracts with median household incomes above the federal poverty line. This is a 139 percent decrease with respect to the sample mean ($p < 0.01$). I cannot reject a null effect of fines and fees for defendants living in census tracts with median household incomes below the poverty line. However, this estimate is noisy and the confidence interval is large (-8.5 p.p, SE=0.193).

In order to examine heterogeneity more systematically, I adapt the method proposed by Chernozhukov et al. (2020) to estimate conditional average treatment effects (CATEs) using generic machine learning inference. I find evidence of significant heterogeneity in treatment effects ($p < 0.01$), which display a pattern similar to the one I find looking at neighborhood income. Fines and court fees seem to have little effect for about 40% of the defendant population, while having very strong deterrent effects for the remaining 60%. The most responsive group is characterized by considerably higher neighborhood income (18% higher; $p < 0.01$) and lower likelihood of a prior involvement with the criminal justice system (46% lower; $p < 0.01$). Additionally, they are also 4 years younger on average ($p < 0.01$), 10% less likely to be a minority ($p < 0.01$), and 16% more likely to have been convicted for a drug offense ($p < 0.01$).

These results point to several mechanisms that could explain my findings. Keeping the expected cost of future offenses constant, being punished may deter a person from reoffending if for instance he/she learns new information about the probability or extent of punishment.[16] I find that fines and court fees reduce reoffending by 13.4 percentage points for individuals

non-financial motive are defined as all crimes that are not categorized as financially motivated.

[15] $25,465 for a family of four with two children in 2018

[16] They may also respond to the salience of the punishment, or via a gift-exchange channel, for example, if they recidivate out of spite ("Someone who is subject to punishment he sees as unjust, may be so embittered that he is more likely to offend than if punishment had been lower. Thus, offenses may increase with punishment over a certain range." (Carr-Hill and Stern, 1979)). I find no evidence of these mechanisms.

with no prior involvement with the criminal justice system (179%; $p < 0.01$), while I see no effects for defendants with a prior record (-5.5 p.p, SE=0.061); this difference is significant ($p < 0.01$). Additionally, the null effects of fines and court fees for defendants from poorer neighborhoods could reflect, for example, less information updating or an offsetting effect through financial distress, as poorer defendants need to engage in crimes that entail economic gains to pay off court fees (Harris, 2016; Natapoff, 2018).[17,18]

Even though I find no evidence of fines and fees driving financially motivated recidivism, fines and fees do seem to be causing some measure of financial distress. I find that monetary sanctions increase reoffending in crimes associated with poverty, such as panhandling or sleeping in public places, by 12.9 percentage points, although this is not significant.

However, the fact that only defendants living in wealthier neighborhoods seem to be responsive to fines and court fees, coupled with evidence of them causing financial distress for the poorest defendants, suggests the need to consider a defendant's financial conditions when levying monetary sanctions, and scaling them appropriately. Since I study the extensive margin, of whether any monetary sanction is imposed at all, scaling sanctions is akin to allocating them only to defendants having incomes above a given threshold. Indeed, marginal treatment effects (MTE) indicate that judges are not targeting fines and court fees in a way that minimizes future reoffending. I explore this possibility by adapting Yadlowsky et al. (2021)'s methods of generic machine learning inference to estimate conditional average treatment effects (CATEs), in order to evaluate alternative ways to allocate monetary sanctions in my sample. I use these CATEs to explore the benefit (relative to a random allocation of fines and court fees in the sample) of using two alternative policies: one that targets the q-th fraction of defendants with the highest CATEs, and one that prioritized individuals living in richer neighborhoods. I find that allocating fines and court fees according to results in an improvement of 40% over allocating fines and court fees randomly, and that half of this improvement (20%) can be achieved by targeting fines and court fees according to neighborhood income alone.

This paper's main contribution is to the literature on monetary sanctions. Observational studies offer mixed evidence, finding both reductions and increases in reoffending, while a set

[17] "[O]ne defendant...recalled that when he had to choose, he paid fines and fees over child support because he thought he was less likely to go to jail that way. He also sold narcotics to pay both debts. As he noted, 'why not do more crimes if you're already in trouble?' " (Nichol and Hunt, 2018).

[18] Note that even fines and fees of low amounts may impact individuals through this income channel, as unpaid legal financial obligations grow over time and other payment penalties accrue. Also, individuals sometimes forego payment of rent or utilities in order to pay their legal financial obligations which may start individuals down a debt spiral eventually making it more appealing to resort to crime. "[To pay off my legal financial obligations] I rob Peter to pay Paul... [Not paying] rent, and car payments, insurance payments..." (Natapoff, 2018).

of papers documents evidence associating monetary sanctions with financial instability and financially motivated recidivism (Beckett et al., 2008; Harris et al., 2010; Harris, 2016).[19] Causal evidence is more limited and the research has focused mostly on non-criminal traffic offenses, finding modest to significant decreases in reoffending rates (Goncalves and Mello, 2017; Dušek and Traxler, 2021; Finlay et al., 2022).[20,21] Two other recent papers study different aspects of the effects of monetary sanctions than my focus in this paper. Pager et al. (2022), randomizes the relief of court debt in Oklahoma County, Oklahoma, find a 0.3 p.p (SE=0.035) reduction on future criminal behavior within one year after the intervention, and a 6 p.p. (SE=0.049) reduction for defendants who are unemployed. However, their treatment is not the levying of monetary sanctions, it is the relief of already incurred court debt; defendants in both treated and control groups likely both receive similar signals about the likelihood and amount of future monetary sanctions (the debt relief wasn't a permanent or universal program). Consequently, this variation isolates the financial distress effects of monetary sanctions but does not provide evidence on the overall effects of monetary sanctions, inclusive of the deterrence effect. Furthermore, the experiment paid off both the court fees and fines from the just incurred offense, but also for previous offenses with outstanding fees and fines, making interpretation of the effects more difficult.

Giles (2022), uses an RD in time design around the date of implementation of an increase in assessed fines in Milwaukee, Wisconsin to study intensive margin variation in the size of fines. He finds that increasing average fines from \$452 to \$745 increases re-offending by 4.19 p.p. within one year after sentencing. This contrasts with my finding of a negative effect of an extensive margin shift in fines and court fees on re-offending. Consistent with my findings, he finds that the effects of fines on re-offending is more positive for lower income defendants. This combination of findings in Giles (2022) and this paper is suggestive that the effects of fines may be non-monotonic, with deterrent effect dominating for the introduction of small fines and the financial distress impact dominating for increases from small to large fines. However, this change in policy was anticipated by the judicial system and there is some

[19] Higher fines are associated with both reductions (Yu, 1994) and increases (Mann et al., 1991) of recidivism for drunk drivers in New York and Ontario. Moffatt and Poynton (2007) reports finding no relationship between fine amount and the likelihood of a defendant reoffending, for a sample of driving offenders in New South Wales. Observational studies focusing not just on driving-related offenses find either a positive (Piquero and Jennings, 2017) or no association between monetary sanctions on reoffending (Gordon and Glaser, 1991).

[20] Goncalves and Mello (2017) and Dušek and Traxler (2021)—using a leniency design and regression discontinuity (RD) and an event study, respectively—study the effect of fines for non-criminal traffic offenses (speeding) and find significant decreases in recidivism.

[21] Finlay et al. (2022) use an RD in time design to study changes in state policy regarding financial sanctions and find that traffic offense fines have null to modest positive effects reducing recidivism and increase short- and medium-run income. They find reductions in recidivism for the smallest fine levels they consider (\$300-\$400).

evidence of changing behavior by prosecutors, judges, and defendants around the change in policy. Specifically, after the increased fines are implemented, the proportion of defendants with property crime histories is 5.3 p.p. ($p = 0.0064$) lower and the proportion of defendants charged for property offenses is 3.8 p.p. lower ($p = 0.0659$), while those charged for drug offenses is 6 p.p higher ($p = 0.0001$).[22] Consequently, some of the estimated effect may reflect this changing behavior or sorting around the policy change.

This paper also adds to the study of specific deterrence—how the experience of punishment affects reoffending. There is limited causal evidence of specific deterrence; it is mostly on incarceration and offers mixed results.[23,24] My set-up allows me to rule out any incapacitation or labor market effects that may confound the analysis, as well as to avoid any criminogenic effects due to criminal networks formed within prison. In this sense, the paper most closely related to this is Dušek and Traxler (2021), who study the effects of speeding tickets and observe reductions in speeding offenses and driving speed. My results extends theirs in terms of the populations under study, as they look at speeding offenders and I study low level misdemeanants.

Finally, a number of papers have studied the graduation of monetary sanctions according to defendants' ability to pay them (McDonald et al., 1992; Turner and Petersilia, 1996; Colgan, 2017). Their focus has been a number of experiments conducted in the late 80s and early 90s in the United States, finding conflicting results and mired with design and compliance issues. I provide new evidence on the effectiveness of scaling sanctions, finding significant decreases in reoffending.[25]

The remainder of the paper is structured as follows. Section 2 provides a brief overview of the criminal justice system in North Carolina and how monetary sanctions and judge assignment operate in this context. Section 3 describes my data and provides some summary statistics. Section 4 describes my identification approach. Section 5 presents the results and studies heterogeneity and mechanisms. Section 6 presents some robustness checks and Section 7 concludes. An Appendix provides additional results and detailed information on administrative information and alternative outcomes used in the analysis.

[22] Giles (2022) doesn't report balance before and after the policy change on defendant income level.

[23] Some find reductions in future criminality (Kuziemko, 2013; Bhuller et al., 2020; Rose and Shem-Tov, 2021), while others find increases (Bayer et al., 2009; Stevenson, 2017; Mueller-Smith, 2015)

[24] There is also sizable literature on the effects of monetary punishment—from regulatory violations—on firms' behavior (see Alm and Shimshack, 2014 for an overview of the literature). In a similar spirit as this paper, Earnhart and Segerson, 2012 find that the effectiveness of increased enforcement—regarding firms' environmental performance—depends on firms' financial status. Increased enforcement can actually lead to worse environmental performance if firms are financially constrained.

[25] These experiments were conducted in Staten Island, New York; Maricopa County, Arizona; Bridgeport, Connecticut; Polk County, Iowa; Josephine, Malheur, and Marion counties, Oregon; Milwaukee, Wisconsin; and Ventura county, California.

2 The North Carolina Criminal Justice System

There are three main challenges to investigating the causal effects of monetary sanctions on criminal reoffending. First, high-quality individual-level data linked to total fines and court fees assessed are difficult to obtain. Second, monetary sanctions are typically correlated to unobservables related to criminality, rendering naive OLS estimates of the relationship between them likely to be biased. And third, monetary sanctions are usually bundled with other penalties, again potentially biasing any causal estimates.

In this section, I outline how the setting and data from North Carolina allow me to address each of these concerns. In sum: administrative data from North Carolina provide detailed information regarding fines, court fees, and reoffending; quasi-random assignment of district court judges allows me to identify causal effects by leveraging a judge leniency design; and finally, reforms enacted in 2014 create a subset of offenses that are liable only to monetary sanctions, letting me study their effect in isolation.

I first discuss the North Carolina criminal justice system and particular context important for my study. I then discuss how fines and court fees are levied in the state. Lastly, I discuss a reform that limited penalties of low-level misdemeanors in North Carolina to fines and court fees alone. I describe the data and sample in Section 3, below.

2.1 District Courts and Judges

I use data on criminal cases prosecuted in North Carolina's District Courts, which process all criminal cases involving misdemeanors and infractions in the state.[26] District Courts are divided into 41 districts across the state and sit in the county seat of each county (see Figure C1 in the Appendix).[27] There are around 270 District Court judges who are elected by voters in their districts to serve terms of four years. Additionally, the Chief Justice of North Carolina's Supreme Court designates one judge per district as chief District Court judge.

The typical timeline of a misdemeanor case is as follows: individuals are typically brought in front of a magistrate judge within 48 hours of their arrest, who sets the conditions for their release. Usually, a defendant is either released without bail (signing a "written promise

[26] The court system in North Carolina consists of two trial court divisions: the Superior Court division and the District Court division (there is also an appellate division, which comprises the Supreme Court and the Court of Appeals—see Appendix Table C1). Superior Courts are the highest of the general trial courts and, broadly speaking, have jurisdiction over all felony cases, as well as over civil cases involving large amounts of money and misdemeanor and infraction cases appealed from a decision in District Courts. District Courts have jurisdiction over civil cases such as divorces and child custody/support lawsuits, as well as all misdemeanors and infractions, unless the misdemeanor was committed as part of the same act as a felony, in which case both are tried together in Superior Court.

[27] They may also preside in certain other cities and towns specifically authorized by the General Assembly.

to appear"), released on bond, or denied pretrial release. The magistrate also determines the date of the defendant's "first appearance" in District Court, which usually takes place within 96 hours of the arrest. In this hearing, the judge explains the charges to the defendant and appoints an attorney to represent them if the charges carry prison sentences or monetary sanctions in excess of $500 and the defendant is deemed "indigent", i.e., cannot afford to retain private counsel. In North Carolina, depending on the jurisdiction, indigent defense is provided either by public defenders—full-time state employees—or by private attorneys who choose to participate in an assigned counsel system.[28]

Criminal trials in District Courts are "bench trials", in which a judge decides the verdict and sentencing, rather than relying on the decision of a jury.[29] However, upon conviction, judges must follow sentencing guidelines which depend on the severity of the crime and the defendant' s previous criminal record. This structured sentencing scheme classifies misdemeanor charges into three classes (from most to least severe: A1, 1, 2, and 3), and defendants into criminal history levels I–III, where higher levels indicate more prior involvement with the criminal justice system.[30] The resulting sentencing grid establishes the sentencing options available to judges: community punishment (probation), intermediate punishment (probation with some additional conditions), and active punishment—incarceration in prison or jail (see Figure C2 in the Appendix).

Notice that Class 3 offenses are only liable to fines for defendants with up to 3 prior convictions. Infractions are also typically punishable by a fine only—and they do not appear on a person's criminal record (examples of infractions include speeding, failure to wear a seat belt, and jaywalking).

Criminal cases are heard in the county court where they are filed. The chief District Court judge creates the schedule of court sessions and assigns judges to preside over them. Then, as per general statute, the district attorney has the responsibility to calendar cases, but in practice this is randomized by the clerk's office (G.S. 7A-61).[31] I discuss this random assignment in more detail in Section 4 below.

[28] Notice that these counsel services are not free of charge to the defendant. In the event of a conviction, the defendant has to reimburse the state for the value of these services, as outlined below in Section 2.2.

[29] Because the Sixth Amendment to the United States Constitution guarantees a person charged with a crime the right to be tried by jury, a defendant convicted of a misdemeanor in District Court has the right to appeal his conviction to the Superior Court for a new trial by jury. A person found responsible of an infraction or a low-level misdemeanor—following the state's 2013 Appropriations Act, discussed below—cannot be sent to jail or prison for the infraction, so the state constitution's requirement of trial by jury does not apply.

[30] N.C. GEN. STAT. §15A-81B, added by 1993 N.C. Sess. Laws 538.

[31] See UNC's School of Government's Prosecutors' Resource Online or Abrams and Fackler, 2018.

2.2 Fines and Court Fees

There are three main categories of monetary sanctions: fines, court costs/fees, and restitution, which differ mainly in their stated purpose. Fines and restitution are intended to punish criminals and compensate victims, while court costs/fees are meant to generate revenue to fund courts and other criminal justice operations.[32,33]

In North Carolina, any person convicted of a crime may be ordered to pay fines or court fees as part of the sentence, and restitution if appropriate.[34] Court costs are broadly divided into basic costs, which apply by default in case of conviction, of around $150, and contingent costs, that are triggered only in certain circumstances, such as being convicted of a specified type of offense, or when certain things happen in the course of the case, such as failing to show up for court, that can go up to $200.[35] All things considered, defendants appearing in District Court can expect to pay a minimum fee amount of $178 for an infraction and $180 for a misdemeanor. See Table 1 for a more detailed breakdown of court costs for the types of cases under study in this paper.[36] Judicial discretion regarding court fees amounts to total or partial waivers, as all court fees are capped at the amounts established by statute.

The amount of fines is at the complete discretion of the judge, except for Class 2 and 3 misdemeanors, whose maximum amount is capped. For Class 2 misdemeanors fines cannot exceed $1,000, and for Class 3 misdemeanors the maximum is $200.[37]

Monetary obligations from criminal and infraction cases are due at the time of conviction, but payment can be delayed to a later date or paid over time for a one-time fee of $20 to cover the State's costs of processing these future payments. If the total amount is not paid within 40 days of conviction (or within 40 days of the date allowed by the court) this triggers what is referred to as a "failure to comply" (FTC), which carries an additional fee of $50 and requires the defendant to appear in court and "show cause", i.e., to explain why they should not be jailed or otherwise penalized for their failure to comply.[38,39] If it is the case that nonpayment was not willful, the court may grant additional time for payment, reduce the owed amount, or revoke the monetary obligation altogether. Finally, if a defendant has defaulted in payment of fines, fees, or other monetary obligations, the court may order that

[32] N.C. Const. Art. IX, §7.

[33] In terms of the allocation of funds, in North Carolina the proceedings of all fines go towards maintaining public schools while court fees finance the court system.

[34] N.C. GEN. STAT. §§15A-1361, 15A-1340.17, 15A-1340.23, 7A-304, and 15A-1340.34.

[35] Court costs are outlined in N.C. GEN. STAT. §7A-304.

[36] See Figures C3 and C4 in the Appendix for the full schedule of court fees for 2019 (it stayed unchanged throughout the study period).

[37] N.C. GEN. STAT. §15A-1340.23.

[38] N.C. GEN. STAT. §§7A-304(a)(6), 15A-1362(c), and 15A-1364.

[39] Because the defendant is at risk of being imprisoned as a result of these hearings, they must be afforded counsel.

unpaid fines or costs be docketed as a civil judgment, a lien on the defendant's real estate.[40]

2.3 2013 Appropriations Act

In North Carolina, as with most of the United States, monetary sanctions are usually levied at the same time as other penalties such as community service, probation, parole, or incarceration. This creates an identification challenge, in the sense that defendants are levied both fines and fees, as well as other punishments simultaneously. To isolate the effects of fines and fees in this paper, I take advantage of the fact that in 2013, in North Carolina, all additional punishments—other than fines and fees—were removed for low level misdemeanors.

To save money on indigent defense expenses—i.e., state-provided defense services— North Carolina made major changes to its criminal justice system as part of the 2013 Appropriations Act.[41] Under this Act, North Carolina's General Assembly enacted a new punishment scheme for the least serious class of misdemeanors (Class 3), limiting punishment to a fine for most defendants. The change applied to offenses committed on or after December 1, 2013. In addition to changing the punishment for Class 3 misdemeanors, the reform reclassified some higher-level misdemeanors (Class 1 and 2) as Class 3 misdemeanors, as well as some Class 3 misdemeanors as infractions.[42] The punishment for offenses reclassified as Class 3 misdemeanors is likewise limited to a fine. Because due to other structured sentencing rules, the maximum fine is usually limited to $200 for Class 3 misdemeanors—which has always been the case for traffic violations—these defendants would not be entitled to indigent defense as per North Carolina provisions this is only the case if they face jail/prison time and/or monetary sanctions in excess of $500, which is the money-saving rationale of the policy.

Therefore, to avoid confounding due to multiple treatments my analysis focuses on all cases started after December 1, 2013.

3 Data and Sample

In this paper, I analyze case-level data provided by the North Carolina Administrative Office of the Courts (NCAOC), which comprise the universe of felony, misdemeanor, and infraction cases filed at District and Superior Courts from 2013 until 2019. I first briefly describe this

[40] N.C. GEN. STAT. §§15A-1365, 7A-455, 7A-455.1, and 15A-1340.38.

[41] Current Operations and Capital Improvements Appropriations Act of 2013, §18B.13, 2013 N.C. Sess. Laws 995, 1303–04. The Joint Conference Committee Report on the Continuation, Expansion, and Capital Budgets, (2013), states that Indigent Defense Services' budget would be reduced by $2,000,000 annually.

[42] *Id.* §§18B.14 and §18B.15, 2013 N.C. Sess. Laws 995, 1304–09, amended by 2013 N.C. Sess. Laws 1594, §§4–6.

data and the specific sample I use for the analysis, and then I provide some descriptive statistics. I provide more details of the NCAOC data in Appendix B.2, while Appendix B.3 describes my cleaning process for this data as well as the analytical sample.

3.1 Data

The NCAOC data include detailed information on sentencing decisions, arrest charges, and characteristics of the defendants. For each case, I identify each judge and defense attorney, and whether the defense attorney is privately retained or providing indigent counsel. This case-level data also includes defendants' home addresses, which I use to obtain census block information from the 2014-2018 American Community Survey.

For the purpose of my analysis, defendants charged with multiple offenses—or recharged for the same crime after a mistrial, in the case of misdemeanors—are collapsed to a single observation. In this scenario, I retain only the earliest filing date, charge characteristics and original sentencing outcomes. When a defendant faces multiple charges I only consider the most severe charge, i.e., the charge associated with the most severe sentencing class.[43]

3.2 Analytical Sample

As mentioned in Section 2.3, I restrict my attention to low level misdemeanors to avoid confounding fines and fees with other punishments. I then make two additional restrictions to the sample to facilitate my analysis.

First, I focus on Class 3 misdemeanors where defendants are subjects to fines only. This represents 23 percent of the universe of non motor vehicle misdemeanors in the study period.[44] Second, I restrict the sample to offenses where defendants cannot waive their right to be present at trial, as waivable cases get handled by magistrates, and not randomly assigned district judges.[45,46] These comprise 14% of remaining cases. Finally, I only consider

[43] This is done since court fees are assessed at the case level, and because for many low-level offenses when more than one charge is made, defendants are fined only for the offense carrying the highest fine (N.C. GEN. STAT. §7A-148). Estimates considering monetary sanctions aggregated to the case level are available upon request.

[44] I consider only non motor vehicle offenses, as non-payment of monetary sanctions in those cases also carries the possibility of driver's license revocation.

[45] The right to be present at trial may be waived in all cases except capital ones (State v. Braswell, 312 N.C. 553, 558 (1985); State v. Daniels, 337 N.C. 243, 256 (1994); State v. Huff, 325 N.C. 1, 29 (1989), vacated on other grounds sub nom., Huff v. North Carolina, 497 U.S. 1021 (1990); ; State v. Hayes, 291 N.C. 293, 296-97 (1976)).

[46] Every year the Conference of Chief District Judges puts out a list of offenses for which magistrates and clerks of court may accept written appearances, waivers of trial or hearing and pleas of guilty or admissions of responsibility, as well as a list of "non-waivable" offenses. I restrict my sample to those cases in the "non-waivable" list.

cases handled by judges who oversee at least 50 cases, and to those cases that are not "singletons" within the set of county-by-time fixed effects.[47,48] In combination, these restrictions leave me with 10,340 cases (44% of non-waivable cases), overseen by 104 judges.[49]

3.3 Descriptive statistics

Table 2 reports descriptive statistics for my analysis sample. The main data set includes 10,340 cases with final dates of disposition between 2013 and 2019. Details on this population's demographic characteristics are provided in Panel A. The sample is predominantly black (57%) and male (73%), overrepresented relative to North Carolina's population (22% and 49%, respectively). The sample is quite young as well, with 67% of defendants below the age of 35, and 42% below the age of 24.

The study population is middle-class on average: based on addresses from the court records, the average defendant lives in a neighborhood in which the median income is $51,000, and where 20% of households earn below the poverty line.

Cases are all Class 3 misdemeanors, the majority of which are drug possession offenses; the rest being property and other, such as liquor laws violations (open container violations, being intoxicated and disruptive in public, etc.). There are no violent charges. 70% of cases in the sample result in a conviction, with almost all of these achieved via pleas.

52% of defendants were levied some monetary sanction. These stem either from fines (29%) or courts fees (50%), with no restitution being levied on any defendant in the sample. Monetary sanctions average $127 for those that receive any, with the breakdown between fines and fees being $14 and $113, respectively (see Figure 1 for the distribution of fines and court fees in the sample).

Finally, it is important to point out that defendants who are levied monetary sanctions are clearly different from those who are not (see Table A3). Defendants that receive monetary sanctions tend to be minorities more frequently, as well as somewhat younger; they also tend to live in neighborhoods with larger median household incomes. Defendants that end up paying any monetary sanction are charged for fewer offenses and are more likely to have a prior criminal record. They are charged for drug offenses more often and less for property offenses. Additionally, they are also less likely to have public representation and more prone to represent themselves. They plead guilty considerably more.

[47] I also drop cases for repeat offenders where the reoffending upgrades the misdemeanor to Class 2 (as is the case of shoplifting under N.C. GEN. STAT. §14-72.1, for example).

[48] I also drop cases of underage purchase of alcohol, pursuant to N.C. GEN. STAT. §18B-302, as driver's license revocation is a possibility.

[49] See Appendix Table A2 for more details on the sample selection.

4 Empirical Strategy

4.1 Overall Approach

I order to estimate the effect of monetary sanctions on future criminal involvement it is important to recognize that unobservables (to us) associated with future criminality that influence judges' decisions regarding monetary sanctions would render OLS regressions biased. I address this endogeneity of monetary sanctions with an IV approach, leveraging the random assignment of cases to district court judges, i.e., using a "judge fixed effects" design.

In judge fixed effects designs it is crucial that the treatment variable of interest be binary (Frandsen et al., 2019) so I consider a dummy variable indicating if the defendant gets levied either a fine or court fees, as this is the relevant margin for defendants: whether they are levied any monetary sanction, any legal financial obligation. Thus, $Any\ LFO = \mathbb{1}_{LFO>0}$. I consider, then, the following model:

$$Y_{ict} = \beta_1 Any\ LFO_{ict} + \beta_2 \mathbf{X}_{ict} + \gamma_{ct} + \varepsilon_{ict} \tag{1}$$

where Y_{ict} is the outcome of interest for individual i in case c and year t, such as future criminality, for example. $\mathbf{X}_{ict}$ is a vector of case- and defendant-level control variables, γ_{ict} are county-by-time fixed effects, and ε_{ict} is an error term.

The key problem for causal inference in this setting is that OLS estimates of equation (1) are likely biased, due to correlation between them having been levied monetary sanctions and unobserved defendant characteristics that are correlated with the outcome. The sign of this correlation, and therefore the bias stemming from it, is unclear. For example, judges may be more likely to fine defendants that display less regret who, in turn, may be more likely to reoffend.[50]

To address this issue, I need variation in whether a defendant gets levied fines or fees orthogonal to unobserved defendant characteristics. The quasi-random assignment of district judges to defendants provides exactly this type of variation. I can use the quasi-randomly-assigned judges as instruments for whether defendants are levied monetary sanctions, with the following first stage in a 2SLS model:

$$Any\ LFO_{ict} = \sum_j \alpha_{1j} \mathbb{1}_{j(i)=j} + \alpha_2 \mathbf{X}_{ict} + \gamma_{ct} + \eta_{ict} \tag{2}$$

[50] At the same time, judges could be inclined to fine defendants they perceive as having a higher ability to pay. Defendants with a higher ability to pay probably have a stronger attachment to the legal labor market, and are thus less likely to reoffend.

with j indexing judges and $j(i)$ indicating judge assignment for defendant i. This specification identifies the local average treatment effect (LATE) for each pair of judges and corresponds to the average treatment effect for individuals who would get a fine or fee if assigned to the stricter judge of the two, but would not if assigned to the more lenient judge.

The problem with this approach is that it is well known that when many instruments are used, 2SLS may be severely biased (Bound et al., 1995). The usual recommendation to address this is to use JIVE or LIML estimators. JIVE—jackknife instrumental variable estimator—, which seems to be the solution the literature has settled on, first suggested by Phillips and Hale (1977) and later by Angrist et al. (1999) and Blomquist and Dahlberg (1999), attempts to improve finite-sample properties by replacing the usual fitted values from the reduced form regression(s) by "omit-one" fitted values which omit observation i when estimating the i-th fitted value, thus eliminating the correlation between the fitted values and the structural equation errors.

Rather than JIVE, it has been suggested that LIML may be a better alternative (Davidson and MacKinnon, 2006), as even though the small-sample bias of JIVE does not depend on the degree of overidentification, it does increase in the number of exogenous covariates present in the second-stage equation (Ackerberg and Devereux, 2009). However, LIML is not consistent with heteroskedastic data (Chao and Swanson, 2004 and Hausman et al., 2012) and, additionally, Kolesár (2013) shows that in the presence of treatment effect heterogeneity the estimand of LIML may even lie outside of the convex hull of LATEs. Kolesár (2013) proposes a modification of the JIVE, the unbiased JIVE (UJIVE) estimator, that remains consistent even in the presence of many instruments and covariates.

In my setting, judge assignment is random only conditional on county-by-time fixed effects, because as-if randomization of cases to judges takes place within county and time periods. Therefore, the construction of the leave-out mean of JIVE requires that first I residualize out county-by-time fixed effects to limit the comparison to defendants at risk of being assigned the same set of judges. Specifically, given the scheduling system in North Carolina, I account for county-by-year-month fixed effects. Because of the large number of fixed effects, I present results using the UJIVE estimator in my main tables.[51]

JIVE and UJIVE are equivalent to 2SLS, in the sense that they initially construct a single instrument, a predictor of treatment status based on the first-stage and, on a second step, they use this constructed instrument as a single instrument to estimate the treatment effect. In the case of UJIVE, that constructed instrument is algebraically equivalent to

[51] I also explore additional estimation strategies, 2SLS using all judge dummies, LIML, JIVE, and Belloni et al. (2014)'s IV Lasso in Appendix C.4.

$$Z_{ict} = \left(\frac{1}{n_{tj(i)} - n_{itj(i)}}\right)\left(\sum_{k=0}^{n_{tj(i)}} Any\ LFO^{*}_{ikt} - \sum_{l=0}^{n_{itj(i)}} Any\ LFO^{*}_{ilt}\right) \tag{3}$$

where $n_{tj(i)}$ is the number of cases assigned to judge j and $n_{itj(i)}$ is the number of cases involving defendant i seen by judge j in year t. $Any\ LFO^{*}_{ict}$ is the residual likelihood of receiving a monetary sanction after removing county-by-time fixed effects.

I estimate the causal effect of having to pay a fine or court fees, then, using this measure of the likelihood of being levied a monetary sanction by a quasi-randomly-assigned judge as an instrument for having any legal financial obligation at the conclusion of a defendant's trial. Essentially, I compare future criminal involvement for defendants assigned to judges with different propensities to levy financial obligations, and interpret the differences as the causal effect of being levied monetary sanctions associated with the difference in fining propensities of the assigned judges.

4.2 Assessing the Instrument

4.2.1 Instrument Relevance

Figure 2 shows the identifying variation in my data: it reports distribution of the residualized judge leniency measure for imposing any monetary sanction within the sample of 104 judges.[52] The median number of cases overseen by a judge is 79 cases and the average is 100 cases. After residualizing out the set of county-by-time effects, the judge leniency measure ranges from-0.925 to 1.070, with a standard deviation of 0.124. Moving from the first to the ninety-ninth percentile of judge leniency increases the likelihood of receiving a fine or a fee by 46 percentage points, an 88% change from the mean likelihood of 52%.

To consider the first-stage relationship between judge assignment and the amount of monetary sanctions, I estimate the following equation for defendant i and case c, assigned to judge $j(i)$ at time t:

$$Any\ LFO_{ict} = \alpha_1 Z_{ict} + \alpha_2 \mathbf{X}_{ict} + \gamma_{ct} + \eta_{ict} \tag{4}$$

where the vector $\mathbf{X}_{ict}$ includes case- and defendant-level covariates, Z_{ict} are the leave-out measures of judge leniency described above, and γ_{ct} are county-by-time fixed effects. Robust

[52] As mentioned in Section 3.3, the variation in the judge leniency measure comes from two sources, as there are no cases warranting restitution in my sample. Judges may choose to levy harsher or more lenient fines, or may choose to waive—all or part of the—court fees. Figure A1 shows the distribution of residualized judge leniency for these two margins separately. In my preferred specification I collapse both of these measures into one, as that captures what I consider the relevant margin for defendants.

standard errors are clustered at the individual and judge level.

Figure 2 also illustrates graphically the first-stage relationship, described in Equation (4), between the residualized measure of judge leniency and whether a defendant gets levied any monetary sanctions, controlling for county-by-time fixed effects. It plots a local linear regression of the likelihood of receiving a monetary sanction against judge leniency, after controlling for county-by-time fixed effects. This likelihood of receiving a monetary sanction is monotonically, and approximately linearly, increasing in the judge leniency measure. A 10 percentage point increase in the residualized judge's fining rate in other cases is associated with an approximately 6.8 percentage points increase in the likelihood of being levied a monetary sanction.

Table 3 presents formal first-stage results from Equation (4). Column 2 begins by reporting results only with county-by-time fixed effects. Column 3 adds baseline case and defendant controls: race, gender, age, the number of charged offenses, indicators for crime type and the type of defense representation, as well as the median household income for the census block group where the defendant resides. The results in Table 3 are consistent with Figure 2: the residualized judge instrument is highly predictive of whether a defendant faces any monetary sanctions. Including controls in column 3 changes very little the magnitude of the estimated first-stage effect, consistent with the quasi-randomness of judge assignment. With all controls my results show that a defendant assigned to a district judge that is 10 percentage points more likely to levy a fine or court fees is 6.8 percentage points more likely to receive a monetary sanction.

The instrument is strong, with first stage F-statistics greater than 120. These F-statistics are Kleibergen-Paap robust F-statistics, which in the just-identified case are equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013). Both of these F-statistics exceed 12.28, which is the critical value they propose for just-identified models to test an IV relative bias of no more than 10% with a significance level of 5.[53]

4.2.2 Conditional Independence

For the instrument to be valid, it must be the case that the residualized measure of judge leniency is orthogonal to unobserved defendant and case characteristics (conditional on court-by-time fixed effects).

I can study this by examining whether observed characteristics of defendants and their cases differ by judge. Table 4 provides strong empirical evidence of the random assignment of cases to district judges within each court in a given time period. The first column of Table

[53] It is not clear how to evaluate many-weak-instrument bias in leniency/examiner designs though (Hull, 2017; Frandsen et al., 2019; Bhuller et al., 2020).

4 tests whether case and defendant characteristics are predictive of whether a defendant receives a monetary sanction. I control for court-by-time fixed effects and cluster the standard errors at the individual and judge level. I find that for each additional concurrent offense defendants are 12 percentage points less likely to be levied any fines or fees, a 26 percent decrease from the baseline mean of 46 percent. Defendants with a prior offense are 12 percentage points more likely to be fined, compared to defendants with no prior offense, a 26 percent increase. Additionally, defendants arrested for property offenses are almost 7 percentage points less likely to be fined than those arrested for other types of offenses, a 15 percent decrease. Finally, individuals who have private representation or that represent themselves are both close to 14 percentage points more likely to be fined compared to those under public representation, a 30 percent increase. Column 2 assesses whether these same case and defendant characteristics are predictive of the judge leniency measure under the same specification. I cannot reject the null hypothesis of all coefficients being zero (joint p-value = 0.37). I find evidence, then, in support of district judges of differing tendencies being assigned very similar defendants.

4.2.3 Monotonicity and Exclusion Restriction

The exclusion restriction requires that judges' leniency imposing monetary sanctions influences defendants' outcomes solely through fines and courts fees. While I cannot directly test this assumption, there is only one other prominent channel that may affect defendant outcomes aside from monetary sanctions, which is the acquisition of a criminal record. Defendants with a misdemeanor record may face higher future punishments if they reoffend, a "general" deterrence effect, or experience limited job prospects, lowering the opportunity costs for future crimes. In such cases, the estimated impact of fines and court fees, without considering conviction, represents a combined effect of these factors.

To address this concern, I first reestimate may main model for defendants charged with drug or property offenses who have prior criminal justice system involvement, as these defendants are always convicted in my sample. The estimates from this analysis, although less precise, align with my main findings (refer to Table A4). Secondly, I follow Mueller-Smith (2015) and incorporate dummy variables for both conviction and the imposition of fines or court fees into my model, instrumenting both with leave-out measures akin to the one described in Equation (3). This leads to the following structural equation:

$$Y_{ict} = \beta_1 Any\ LFO_{ict} + \beta_2 Conviction_{ict} + \beta_3 \mathbf{X}_{ict} + \gamma_{ct} + \varepsilon_{ict} \tag{5}$$

and first stage:

$$Any\ LFO_{ict} = \alpha_1 Z^{LFO}_{ict} + \alpha_2 Z^{Conv}_{ict} + \alpha_3 \mathbf{X}_{ict} + \theta_{ct} + \eta_{ict} \tag{6}$$

$$Conviction_{ict} = \delta_1 Z^{LFO}_{ict} + \delta_2 Z^{Conv}_{ict} + \delta_3 \mathbf{X}_{ict} + \kappa_{ct} + \eta_{ict} \tag{7}$$

where Z^{LFO}_{ict} and Z^{Conv}_{ict} are judge $j(i)$'s leave-out mean rates of levying monetary sanctions and conviction, respectively. These rates serve as instruments for defendant i's imposition of monetary sanctions and guilty verdict in court. This model has two endogenous sentencing variables which are simultaneously instrumented, identifying the effects of fines and court fees using residual variation after accounting for judges' tendencies to convict. Table A5 reports estimates from this model, which are again consistent with my main results, providing further evidence of the exclusion restriction being satisfied.[54]

Additionally, the impact of judge assignment on monetary sanctions needs to be monotonic across defendants for me to be able to interpret my estimates as a well-defined LATE. This monotonicity assumption implies that defendants who are spared any fines by stricter judges would also be spared by more lenient judges, and that defendants levied fines by more lenient judges would be levied fines as well by stricter judges.

Regarding monotonicity, Frandsen et al. (2019)—FLL hereafter—propose a test for the joint null hypothesis that both the exclusion and monotonicity assumptions hold. Their result states that if conditions for LATE identification are satisfied, outcomes averaged at the judge level will be a continuous function, with bounded slope, of the judge propensity to treat, and the test consists of examining whether observed outcomes averaged by judge are consistent with such a function. I summarize the process below and provide full details in Appendix Section C.3.

In order to do this you first regress the outcome on a flexible function of the judge propensity. FLL use splines to represent this function, which are essentially piecewise polynomial segments strung together at "knots". For a given polynomial order, then, the number of knots determines the function's flexibility.

Second, you regress the residuals from the previous step on judge indicators and test whether the coefficients are jointly zero and whether the slopes of the function are within the bounds dictated by the support of the outcome. Finally, you combine the fit component and slope component of the test via a weighted Bonferroni procedure to produce a single joint test.

Table A6 shows that I fail to reject this null for various numbers of knots—using a

[54] It is important to note, however, that the interpretation of these estimates is not straightforward in an heterogeneous treatment effects framework, as the complier population is not readily identifiable.

quadratic spline, as suggested in Frandsen (2020)—and Bonferroni weights. It is important to note, however, that even though I cannot reject the null there is a clear pattern of decreasing p-values as I increase both the weight on the fit component of the test as well as the number of knots. This is concerning since it could be the case that the rejections are mainly due to a lack of power (with a large number of judges, the slope component of the test has little power).

A weaker testable implication of the monotonicity assumption is that first-stage estimates should be positive for any subsample (Dobbie et al., 2018; Bhuller et al., 2020). Tables A7 and A8 present these first-stage results separately by type of offense, type of representation, whether defendants have a prior record, by defendant race, and by income quintile (using the full sample of cases to calculate judge leniency). I find that the residualized measure of judge leniency is consistently positive and sizable in all subsamples. FLL show that this actually tests an average monotonicity assumption, which implies that the covariance between a defendant's judge-specific treatment and the judge's overall propensity to levy fines or court fees has to be weakly positive: judges who are more lenient overall should be more likely to not levy fines or court fees in any observable subgroup.

4.3 Treatment Effect Heterogeneity

It is a main concern establishing whether, beyond the average, monetary sanctions impact poorer defendants differentially. Further, treatment effect heterogeneity across other margins, such as prior involvement with the criminal justice system and defendant age, allows me to tease out different mechanisms driving any response.

My approach follows three steps. First, I characterize how estimated treatment effects vary across populations with different neighborhood incomes, prior involvement with the criminal justice system, and defendant age. To extend this analysis, I employ generic machine learning methods described in Chernozhukov et al. (2020)—CDDF hereafter—to estimate conditional average treatment effects (CATEs). In this way, I fully describe treatment effects heterogeneity and characterize those defendants most and least responsive to monetary sanctions.

I then estimate marginal treatment effects (MTEs) to explore unobserved heterogeneity in treatment effects, as indexed by a judge's latent propensity to levy fines or court fees, and assess whether judges are currently targeting monetary sanctions optimally, with respect to future reoffending. Finally, using the already estimated CATEs, I follow Yadlowsky et al. (2021)—YFSBW henceforth—to impute counterfactual responses and compare alternative allocations of fines and court fees.

4.3.1 Conditional Average Treatment Effects

CDDF's approach requires a binary treatment that's randomly assigned.[55] I define treatment as having been assigned a judge with an above-median propensity to levy a fine or court fee.[56] Then, I randomly partition the data evenly into an "auxiliary" sample and a "main" sample. This sample splitting approach limits potential overfitting, due to the proxy predictors and the parameters of interest being estimated using the same observations, and relaxes the conditions necessary for consistent estimation.

Next, I predict recidivism, $\hat{Y}_{ict}$, in the auxiliary sample for treated and control units using an elastic net algorithm (Zou and Hastie, 2005) that incorporates all covariates, denoted Z_{ict}, from Equation (1).[57] This gives me two prediction models, $\hat{\mu}_1(\cdot)$ and $\hat{\mu}_0(\cdot)$, where

$$\mu_1(Z) \coloneqq \mathbb{E}[Y(1) \mid Z] \text{ and } \mu_0(Z) \coloneqq \mathbb{E}[Y(0) \mid Z].$$

Their difference, $S(Z_{ict}) = \hat{\mu}_1(Z_{ict}) - \hat{\mu}_0(Z_{ict})$, represents the change in a defendant's predicted likelihood of recidivating due to being levied a fine or a court fee, a proxy predictor of the (true) conditional average treatment effect (CATE), $s_0(Z_{ict})$. CDDF's idea is that although this predictor is most likely biased, one can still use it to conduct valid inference regarding heterogeneity.[58] I also estimate the propensity score, $\hat{p}(Z_{ict})$, in the auxiliary sample as well, in the same way I estimated predicted recidivism.

Now, having constructed the proxy predictor, $S(Z_{ict})$, I estimate the following weighted regression in the main sample:

$$\begin{aligned} Y_{ict} = \alpha_1 + \alpha_2 \hat{Y}^0_{ict} + \beta_1(Any\ LFO_{ict} - \hat{p}(Z_{ict})) + \\ \beta_2(Any\ LFO_{ict} - \hat{p}(Z_{ict}))(S(Z_{ict}) - \bar{S}) + \varepsilon_{ict}, \quad (8) \end{aligned}$$

where $\bar{S}$ is the average of $S(Z_{ict})$ in the estimation sample and the weights are equal to

$$w(Z_{ict}) = \frac{1}{\hat{p}(Z_{ict})(1 - \hat{p}(Z_{ict}))}$$

To reduce noise, I drop observations with propensity scores below 0.01 or above 0.99.[59]

[55] I summarize the estimation process below and provide full details in Appendix Section C.2.

[56] In the same spirit as Deryugina et al. (2019).

[57] I choose the elastic net as it produces the proxy predictor most correlated with the true conditional average treatment effect function (compared to boosting, support vector machine, and random forest—see Table A17), following the procedure described in Section C.2.1

[58] This is effectively what is called a T-learner, an estimator of the CATE constructed from two separate estimators, one for each regression function defined by treatment status. See Künzel et al. (2019) for specific reasons why this particular CATE estimator may be biased.

[59] I also estimate alternative specifications, trimming the top and bottom 5 percent of observations and

CDDF show that, under general conditions, $\beta_1 + \beta_2(S(Z_{ict}) - \bar{S})$ is the best linear predictor of the CATE, with $\mathbb{E}[\hat{\beta}_1]$ an unbiased estimate of the average treatment effect, $\mathbb{E}[s_0(Z_{iCt})]$. In addition, rejecting the null hypothesis that $\beta_2 = 0$ implies the presence of heterogeneity and the relevance of $S(Z_{ict})$ as a predictor of the CATE.

Next, I estimate group average treatment effects, $\mathbb{E}[s_0(Z_{ict}) \mid G]$, where the groups G are defined as disjoint intervals of $S(Z_{ict})$. Specifically, I estimate the following weighted regression:

$$Y_{ict} = \alpha_1 + \alpha_2 \hat{Y}^0_{ict} + \sum_{i=1}^{K} \gamma_k (Any\ LFO_{ict} - \hat{p}(Z_{ict})) \mathbb{1}_{S(Z_{ict}) \in G_k} + \varepsilon_{ict}, \tag{9}$$

where the indicator $\mathbb{1}_{S(Z_{ict}) \in G_k}$ is 1 if $S(Z_{ict})$ lies in the kth interval and 0 otherwise. The weights are the same as in Equation (8). The coefficients of interest, λ_k, correspond to the average treatment effect for each group k, where the groups are defined by quintiles of the $S(Z_{ict})$ distribution.

4.3.2 Marginal Treatment Effects

I also explore heterogeneity by examining MTEs. Ignoring subscripts for simplicity, I model the observed outcome as $Y = I \times Y(1) + (1 - I) \times Y(0)$, where I is an indicator for treatment (being levied a fine or court fee) and $Y(1)$ and $Y(0)$ are the associated potential outcomes, which are a linear function of both observable (X) and unobservable factors. The choice of treatment by a judge is given by $I = \mathbb{1}_{v(X,Z)-U}$, where v is an unknown function, U is an unobserved continuous random variable, and Z is my judge stringency instrument. One can normalize the distribution of $U \mid X = x$ to be uniformly distributed over $[0, 1]$ for every value of X. Under this normalization, it is straightforward to show that $v(X, Z)$ is equal to the propensity score $p(X, Z) \equiv \mathbb{P}[Y(1) - Y(0) \mid X = x, Z = z]$.

The MTE is defined as $\mathbb{E}[Y(1) - Y(0) \mid U = u, X = x]$. The dependence of the MTE on U for a fixed X reflects unobserved heterogeneity in treatment effects, as indexed by a judge's latent propensity to choose to levy a fine or court fee for a defendant (where U captures unobserved characteristics of the defendant, which influence the judge's choice). The choice equation implies that, given X, defendants with lower values of U are more likely to take treatment, regardless of their realization of Z. Following Brinch et al. (2017), I assume separability between observed and unobserved heterogeneity in the treatment effects which, together with the assumption of an exogenous instrument that satisfies monotonicity is sufficient to allow point identification of the MTEs over the unconditional support of the

using a Horvitz-Thompson transformation, as in Chernozhukov et al. (2020).

propensity score.

4.4 Rank-Weighted Average Treatment Effects

The presence of treatment effect heterogeneity motivates interest in the question of how to best allocate fines and court fees. Is there a better way to assign treatment? Should fines and court fees be tailored to specific defendants?

In the simplified binary-treatment setting I described above this would translate to whether it would be preferable to waive fines and court fees for certain individuals (and not others), and what would be the best way of making this choice.

In order to study this, I follow YFSBW and study different treatment allocation rules. In particular, I want to explore how a random allocation compares to an allocation prioritizing those defendants that are most likely to be deterred by fines and court fees, and an allocation that prioritizes targeting defendants living in wealthier neighborhoods. I summarize YFSBW's method below and provide more details in Appendix Section C.2.2.[60]

I will compare 2 different allocation rules: an allocation prioritizing those defendants expected to have the largest deterrence effects from the experience of receiving a fine or a court fee and an allocation according to defendants' neighborhood income.

In order to identify those individuals most likely to be deterred by a fine or court fees, I first estimate CATEs following the same process as in section 4.3, by building two models, one for treated and one for control units, and estimating the CATE as their difference. After this, then, I have 2 different scoring functions: one that ranks higher those defendants who are expected to be the most deterred by fines and court fees, $S_{CATE}(X_i)$; and one that prioritizes wealthier defendants (according to their neighborhood), $S_{inc}(X_i)$.

Now, I split the sample according to deciles of each of these scoring rules, and calculate for each of them what YFSBW dub the Targeting Operator Characteristic (TOC), which compares the groups defined by the scoring function to the overall ATE. More specifically, I estimate the TOC by comparing sample averages of the CATE estimator:

$$\hat{\text{TOC}}(q; S) = \frac{1}{\lfloor qn \rfloor} \sum_{j=1}^{\lfloor qn \rfloor} \hat{\Gamma}_{j(i)} - \frac{1}{n} \sum_{j=1}^{n} \hat{\Gamma}_i, \tag{10}$$

where $\mathbb{E}[\hat{\Gamma}_i \mid X_i] \approx \mathbb{E}[Y_I(1) - Y_i(0) \mid X_i]$. The RATE estimator is the sample average of the different TOCs:

$$\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} \hat{\text{TOC}}(\frac{j}{n}; S) \tag{11}$$

[60] In practice I use the function `rank_average_treatment_effect` from the package `grf`.

where S denotes the different scoring rules.[61]

5 Results

In this section, I present my main results on the effects of fines and court fees on the likelihood of defendants reoffending, and distinguish by whether it is financially or non-financially motivated. I characterize the role of compliers in my 2SLS estimates. I then explore the heterogeneity of these results, in particular with respect to income. I also assess the effectiveness of the current allocation of fines and court fees, and explore the effects of alternative allocations. Finally, I explore some possible mechanisms.

5.1 Main Effects

Table 5 presents OLS and UJIVE estimates of the overall impact of fines and fees on future criminal involvement (Panel A), and distinguishing by whether it is financially (Panel B) or non-financially (Panel C) motivated. Columns (1)–(3) report OLS estimates. Column (1) reports results only with county-by-time fixed effects. Column (2) adds baseline case and defendant controls. Columns (4) and (5) report UJIVE estimates where I instrument for whether the defendant was levied any fines or court fees using the leave-out measure of judge leniency described in Section 4, with (column (4)) and without (column (5)) baseline controls. All standard errors are clustered at the individual- and judge-level.

Panel A estimates show a strong negative association of 9 percentage points between monetary sanctions and any reoffending ($p < 0.05$). In Figure 3 I see that this is driven mainly by a reduction in felony reoffending, in particular property offenses.

To study this pattern in more detail, in Panels B and C of Table 5 I estimate the effect of monetary sanctions on the probability of financially motivated reoffending and the probability of non-financially motivated reoffending. I find the effects are driven exclusively by reoffending for financially motivated crimes.

In Panel B I observe that for all specifications, fines and fees are negatively associated with financially motivated reoffending. The magnitudes of the OLS estimates are moderately

[61] More generally, the RATE estimator is defined as $\hat{\theta}_{\alpha(s)} = \frac{1}{n}\sum_{j=1}^{n} \alpha\left(\frac{j}{n}\right) \hat{\text{TOC}}(\frac{j}{n}; S)$ where $\alpha(\cdot)$ is a weight function. Implicitly, I am using a constant weight function, which implies that the RATE estimator coincides with the area under the TOC curve (see Figure A4), and following YFSBW I refer to this estimator as AUTOC(S). The AUTOC places more weight on areas under the curve where the expected treatment benefit is largest. When non-zero treatment effects are concentrated among a small subset of the population this weighting is powerful for testing against a sharp null (AUTOC = 0). I also report results using linear weights, $\alpha(q) = q$, which weigh the TOC by the fraction treated, placing as much weight on units with low treatment effects as on units with high treatment effects. When non-zero treatment effects are more diffuse across the entire population, this weighting tends to give greater power when testing against a null effect.

sensitive to the addition of baseline case and defendant controls: in OLS results with only county-by-time fixed effects I find that a defendant who is levied a fine or a fee is 5.4 percentage points less likely to reoffend, a 60.7 percent decrease from the sample mean ($p < 0.01$; column (1)). When I add case and defendant controls, the estimate drops to 3.5 percentage points ($p < 0.01$; column (2)).

Estimates in column (2) could still be biased if judges are choosing on average to levy fines or fees on defendants who have lower risks of reoffending, despite controlling for a rich set of county-by-time fixed effects and case and defendant controls as well. It is for this reason that my identification strategy involves using the quasi-randomly assigned judges as instruments for whether defendants get levied any fines or fees.

The 2SLS estimates with controls in Panel B, column (5)—my preferred specification—indicate that the levying of fines or fees reduces reoffending within two years by 10.3 percentage points ($p < 0.01$), a 116 percent decrease from the sample mean (224 percent decrease relative to the mean for compliers). The direction of the selection bias then was positive: judges on average levy fines and fees on defendants more likely to recidivate.

Panel C reports estimates for the effects of monetary sanctions on non-financially motivated reoffending. OLS results in Panel C paint the opposite picture of Panel B: fines and fees seem to increase non-financially motivated reoffending. Estimates from columns (1) and (3) indicate that defendants are roughly 3.3 percentage points more likely to reoffend when levied fines or fees ($p < 0.01$). However, UJIVE estimates from column (5) show that after controlling for selection bias there is no relationship between fines and fees and non-financially motivated recidivism.

Figure 4 shows how the effect of monetary sanctions evolves over time. For financially motivated reoffending I see a gradual but sustained drop in the likelihood of reoffending until around 18 months, when the effects plateau. There seem to be no effects for non-financially motivated offenses.

5.2 Understanding the LATE

The IV estimates represent the LATE for defendants who would have received a different fining decision had their case been assigned to a different judge, the "compliers." To better understand this LATE, I characterize the number of compliers and their characteristics following the approach developed by Abadie (2003) and extended by Dahl et al. (2014) and Frandsen et al. (2019).[62] See Appendix C.1 for a more detailed description of these

[62]Frandsen et al. (2019) generalize Abadie (2003) and show that the complier mean for a characteristic can be recovered through a regression of the treatment interacted with that characteristic on the treatment instrumented with the judge IV.

calculations.

In Table A9 I estimate these shares and find that the complier share is approximately 54 percent, thus the IV estimates are relevant for a large share of the sampled population. 29 percent of the sample are "never takers," and 17 percent are "always takers."

Individual compliers cannot be identified, but it is possible to describe their observable characteristics following Frandsen et al. (2019). I do this in Table A10. Compliers in the sample are 5 percentage points less likely to be charged with more than one offense, 11 percentage points less likely to be charged with a property offense, 7 percentage points more likely to be charged with a drug offense, 12 percentage points less likely to be represented by a private attorney, 21 percentage points more likely to represent themselves in court, 6 percentage points less likely to be younger than 25 years old, 3 percentage points more likely to be between 25 and 34 years old, and 10 percent less likely to be in the first quartile of household income and 12 percentage points more likely to be in the fourth, compared to the average defendant.[63] Compliers are not systematically different from the average defendant by race or gender, however.

It is important to remember then that IV estimates represent an average treatment effect for the complier population and that therefore, besides selection bias, another possible explanation for differences between IV and OLS estimates is effect heterogeneity, in that the average causal effect for compliers may differ to the mean impact for the whole sample. In order to explore this possibility, it is useful to characterize compliers by their observable characteristics. I follow a procedure similar to that developed by Bhuller et al. (2020). Specifically, I begin by splitting the sample into 10 mutually exclusive and collectively exhaustive subgroups based on the predicted probability of being levied any fines or fees. This predicted probability is a composite index of all the observable characteristic. I then estimate the first-stage equation (3) separately for each subgroup, calculating the proportion of compliers in each. I then reweight the estimation sample so that the proportion of compliers in a given subgroup matches the share of the estimation sample for that subgroup. Reweighted OLS estimates are presented in column (3) of Table 5. Results suggest that the differences between the UJIVE and OLS estimates cannot be accounted for by heterogeneity in effects, at least due to observables.

5.3 Treatment Effect Heterogeneity

In this section, I first analyze whether there are differential impacts of fines and court fees according to defendants' neighborhood income, and explore if there are any additional sources

[63]Where household income is proxied with ACS' median household income in the census tract where they reside.

of heterogeneity. Then, notice that in the presence of treatment effect heterogeneity, the random assignment of judges that differ in leniency allows for the possibility of suboptimal treatment allocations. I explore this by estimating marginal treatment effects (MTEs) and studying the relationship between the MTEs and defendants' treatment propensities (under the assumption that reductions in future reoffending are the main purpose of monetary sanctions). I then assess whether judges are levying monetary sanctions to those defendants that would be most deterred from incurring in future crime, and calculate the effects of alternative counterfactual allocations of fines and court fees, in particular levying fines and court fees to those defendants that are most likely to be able to afford them.

To assess whether an income channel is driving poorer defendants toward crime I estimate Equation 1 separately within subsamples defined by quartiles of defendants' neighborhood income.[64] I find strong negative effects for defendants living in wealthier neighborhoods (see Figure 5).[65] Defendants living in census tracts in the third and fourth quartiles of median household income are 13.4 and 15 percentage points less likely to engage in financially motivated crime 2 years following their disposition. This corresponds to 179 ($p < 0.01$) and 224 ($p < 0.05$) percent decreases with respect to their respective sample means. I find no effects for defendants living in the first and second quartiles (0.012, SE=0.081; -0.094, SE=0.064).

Because the first quartile is relatively high, $33,000, in Table A12 I also split the sample by whether defendants' census tract median household income lies above or below the federal poverty line for North Carolina.[66] I find that fines and fees reduce the likelihood of reoffending by 12.3 percentage points for defendants living in census tracts with median household incomes above the federal poverty line. This is a 162 percent decrease with respect to the sample mean ($p < 0.01$). I cannot reject a null effect of fines and fees for defendants living in census tracts with median household incomes below the poverty line. However, this estimate is noisy and the confidence interval is large. The 90 percent confidence interval goes from -37.6 percentage points to 30.2 percentage points, compared to a sample mean of 0.130.

There could be factors that affect the likelihood to reoffend on a financially motivated offense even after conditioning on (neighborhood) income. I follow CCDF's approach to investigate this possibility. Table 6 presents estimates of β_1 and β_2 from Equation (8), the intercept and slope of the best linear predictor of the CATE. My preferred specification in

[64]One limitation of the data is that it does not include any information regarding defendants' income. I deal with this by matching defendants' addresses to census tracts and then proxying their income with American Community Survey estimates of the median household income within the census tract where they reside.

[65]Table A11 reports point estimates and standard errors.

[66]This was $25,465 for a family of four with two children in 2018, for example.

column (1) indicates that being assigned a harsh judge decreases the likelihood of recidivating committing a financially motivated offense by 7.8%. I also reject the null hypothesis of no heterogeneity ($p < 0.01$). I obtain similar results using the Horvitz-Thompson transformation (column (2)) and when increasing the trimming threshold to 5 percent, transformed or otherwise (columns (3) and (4)).

In Figure 6 I show how the average treatment effect varies across groups of observations with different predicted conditional average treatment effects, $S(Z_{ict})$. I find small and statistically insignificant treatment effects for defendants above the third quintile of the $S(Z_{ict})$ distribution. The treatment effect for those falling in the third quintiles is slightly larger, while the estimated treatment effects becomes much larger for observations in the bottom 2 quintiles of the distribution, with estimated decreases in the likelihood of financially motivated recidivism of almost 9%.

Notice that the pattern of heterogeneous treatment effects displayed in Figure 6, where observations are grouped according to their predicted conditional average treatment effect, is similar to the pattern of heterogeneity shown in Figure 5, where defendants are grouped according to their neighborhood income. Table 7 provides further evidence on this point by comparing the characteristics of those in the top 20 percent of the $S(Z_{ict})$ distribution to those in the bottom 60 percent. Notoriously, defendants in the bottom 60 percent are 50 percent less likely to have a prior record and live in neighborhoods with $9,000 lower median incomes than those in the top 20 percent. Overall, my results suggest that neighborhood income is also a good proxy for vulnerability to fines and court fees.

5.3.1 Marginal Treatment Effects

Marginal treatment effects (MTEs) are defined as the treatment effect at a particular value of a defendant's (unobservable) propensity to receive a fine or a court fee, where this propensity captures defendants' unobserved characteristics, which influence the judge's choice. Intuitively then, because MTEs trace the causal effect of receiving a monetary sanction along these defendants' treatment propensity, they allow you to assess treatment effects for defendants that are being treated and those that are not.

Figure 7 displays the estimated MTEs. These show a downward slope throughout, with point estimates above zero only in the bottom decile of the predicted unobserved resistance to levy fines or court fees. Marginal defendants who are closer to never-takers in the IV framework experience the largest decreases in recidivism when levied monetary sanctions. This is evidence of judges not targeting fines and court fees to those defendants that would be most disincentivized from future reoffending.

As shown by Heckman and Vytlacil (1999, 2005, 2007) conventional treatment effect parameters–such as the (overall) average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treatment effect on the untreated (ATUT)–can be expressed as weighted averages of the MTEs. Recovering these treatment parameters for the entire population, however, requires full support of the propensity score on the unit interval. Figure 8 graphs the propensity score distributions for the treated and untreated samples, where the dashed lines indicate the upper and the lower points of the propensity score with common support (after trimming 1% of the sample with overlap). Since I do not have full support, I follow Carneiro et al. (2011) and rescale the weights so that they integrate to 1 over the region of common support. I report rescaled estimates of the ATE, ATT, and ATUT in Table 8, where these weighted averages are obtained by integrating the MTEs over the propensity score/unobserved resistance to treatment for the relevant sample. These estimates are similar in magnitude to the LATE, although the recidivism effects of monetary sanctions seem to be larger for the treated population.

It is important to note that the estimation and interpretation of MTEs requires a strict monotonicity assumption, as well as additive separability between observed and unobserved heterogeneity in the treatment effects (Brinch et al. (2017)), stronger assumptions than those required to estimate the LATE. And, although I could not reject Frandsen et al. (2019)'s joint null of both strict monotonicity and the exclusion restriction holding, as I mentioned in Section 4.2.3 the pattern I see in Table A6 I interpret these estimates with caution.

All these estimates are based on a local IV approach using a global cubic polynomial specification. In Table 8 I probe the stability of these estimates to various specifications of the empirical model. Reassuringly, the estimates based on linear and quadratic specifications yield similar estimates, as does a semiparametric specification based on local linear regressions.

5.3.2 Rank-Weighted Average Treatment Effects

In Section 4.4 I defined two alternative allocations, according to two different scoring rules, the conditional average treatment effect (CATE) itself and neighborhood income. That is, these rules score higher defendants with larger treatment effects and neighborhood income, respectively, and define two alternative treatment allocations, one prioritizing defendants that would be most deterred from a monetary sanction, and another prioritizing wealthier defendants.

I then estimate, for each scoring rule, the Targeting Operator Characteristic (TOC) for all deciles of the rule, which is just the difference between the average treatment effects for those individuals above each decile and the overall ATE (ITT in my case), where counterfactual treatment effects are imputed using the estimated CATEs. Finally, to obtain the Rank-Weighted Average Treatment Effect (RATE) I just average all the TOCs, weighting it by the fraction treated for the case of the Qini coefficient.

For each prioritization rule, I estimate a 95% bootstrapped confidence interval for the RATE using 10,000 bootstrapped samples (from the main sample in the case of the CATE rule); and then calculate the p-value of a two-sided test of the null hypothesis that the RATE is zero as described in Section C.2.2.

Figures 9 and 10 show the TOC curves for each of the prioritization rules for financially and non-financially motivated reoffending, and Table 9 shows the RATE metrics (AUTOC and Qini coefficient) summarizing these curves, along with their confidence intervals.

Consistent with the results from Section 5.3 I find robust evidence of heterogeneous treatment effects of fines and court fees on financially motivated recidivism, under both the AUTOC and Qini metrics. On the other hand I find no significant evidence of heterogeneous treatment effects for non financially motivated reoffending. Further, in terms of financially motivated reoffending, prioritizing according to neighborhood income improves over a random allocation by either 1.8 percentage points ($p < 0.05$) or 0.4 percentage points ($p < 0.1$), depending on the metric, half the improvement from an optimal allocation, which is 3.6 percentage points according to the AUTOC ($p < 0.01$) or 1.1 percentage points ($p < 0.01$) according to the Qini. Notice that 3.6 p.p. corresponds to a 35% improvement over the LATE estimated in Table 5.

Looking at the TOC curves in Figure 9 suggests why estimates for the Qini coefficient are smaller, as in these curves, we can see that the average treatment effect among the highest prioritized individuals according to either rule is much higher than for the rest, and it tapers fast, especially for the neighborhood income rule. The Qini coefficient weights the TOC by the fraction treated, so it downweights the importance of these small groups with large treatment benefit, which are more influential in the AUTOC.

Interestingly, for non financially motivated reoffending both metrics are not significantly different to random prioritization.

5.4 Mechanisms

The results of the previous section are consistent with individuals that, upon receiving a fine and/or having to pay for court fees, revise upwards their expectations of either the likelihood or the amount of future monetary punishment. Higher expected costs then deter them from reoffending in financially motivated crimes. It is also possible that defendants' priors regarding monetary sanctions are accurate, but that their experience makes them more salient. It is possible as well that a "spite" channel is also in effect, but that gets overshadowed by these previous two. Finally, these are average effects and could be masking important heterogeneity across income, as the income channel is associated with reoffending only for those poorest defendants.

5.4.1 Types of Offenses and Timing

Results in Table 5 are consistent with a deterrence story, where defendants have inaccurate information regarding monetary sanctions—either about their likelihood or their amounts. Deterrence stems from this information being revised upwards. And this deterrence is only for financially motivated offenses, as that is the relevant margin of offenses.

Through the gift-exchange channel I would expect increased recidivism due to fines or fees for all types of offenses. In Table 5 the estimate for financially motivated offenses is significantly negative ($p < 0.05$), while the estimate for non-financially motivated offenses is small and nonsignificant (0.010, SE = 0.044). Additionally, in Figure 3 I see no effects of fines and fees for any of the broad crime categories, except for property offenses where they have a negative effect (-0.038; $p < 0.1$). This is still consistent with the deterrence channel, as most property offenses are for economic gain.

Finally, my main results use a two-year follow-up period for all defendants when considering the effects of fines and fees on reoffending. Figure 4 shows the evolution over time of these effects. The figure presents UJIVE estimates of Equation 1 similar to column (5) of Table 5 but for one-month increments in the follow-up period. These results are inconsistent with a salience argument. There is a steady drop in the likelihood of a new financially motivated criminal involvement within the first year which then remains steady throughout. If salience were the driving mechanism we would see, if anything, the opposite temporal pattern.

5.4.2 Prior Involvement with the Criminal Justice System

One way to investigate if it is inaccurate information regarding monetary sanctions what is driving the effects in Table 5 is to study the effects of monetary sanctions separately for defendants with and without experience in the criminal justice system. Defendants that have prior experience probably have better information regarding monetary sanctions, having experienced them before, and should not respond to monetary sanctions due to a deterrence channel.

In Table A15 I report estimates of the effects of monetary sanctions separately for samples of individuals with and without prior experience in the criminal justice system. Defendants with no prior knowledge of the criminal justice system are the ones that are driving the results. Upon experiencing a fine or having to pay court fees these defendants are 13.4 percentage points less likely to engage in financially motivated reoffending, a 178 percent decrease from the mean ($p < 0.01$). For defendants without any prior experience with the criminal justice system the effects are not significant (-0.055, SE = 0.061).

5.4.3 Poverty crimes

Since I cannot rule out differential deterrence across income, I cannot assume an income channel criminalizing poorer defendants. One way to gain some insight into this channel is to study heterogeneity in regard to different types of financially motivated offenses.

In the US there are certain behaviors that are criminalized that would not fit the usual standard for crimes. For example, what are dubbed "Vagrancy" crimes according to the FBI Uniform Crime Reports, such as panhandling or sleeping in public places, or as driving with a license revoked due to the nonpayment of fines or court fees.[67] It is important to distinguish if these are the types of offenses driving the income heterogeneity we see in Figure 5, as this type of offenses, if positively encouraged due to fines or court fees, most likely signal financial distress. For this reason I will refer to this type of offenses as "poverty" offenses.

Results from Figure 11 show that fines and fees do indeed drive up these poverty offenses among poorer defendants.[68] For defendants living in census tracts with median household incomes below the poverty line I find that fines and fees increase the likelihood of reoffending in these "poverty" offenses by 12.9 percentage points, although not statistically significant. This is a 253 percent increase with respect to the sample mean. For wealthier defendants the effect is almost zero and nonsignificant as well (0.001, SE=0.024).

6 Robustness Checks

In this section I explore possible threats to both the causal and conceptual interpretations of my results. First, I check whether the heterogeneity by income described before is driven by treatment effect heterogeneity along margins other than income, but that covary with it. This could happen, for instance, if it is private representation what is driving the effects, and more defendants are privately represented at higher income quartiles. Second, I check if it is treatment "intensity" what varies with income, in the sense that judges levy higher fees and court fees in districts that comprise census tracts with higher median incomes. Third, I also check the robustness of my results to alternative model specifications, since the baseline outcome means are very low and a linear probability model may be misspecified in such situations. Fourth, I study alternatives to the UJIVE to identify the effects of fees and fines.

[67] This is the most common type of reoffending offense in my sample. See Figure A3.

[68] Point estimates and standard errors are reported in Table A14.

Finally, I assess the external validity of these results using a natural experiment that took place in mid 2018.

6.1 Complier Population Heterogeneity

An alternative explanation for the results in Figure 5 is treatment effect heterogeneity. After all, instrumental variables estimations recover "local" average treatment effects, as in for the complier population. When analyzing heterogeneity by income quartiles I am effectively looking at different populations and, therefore, possibly different complier populations as well. If treatment effects vary across these different population, it could be the case that this is what is driving the heterogeneity.

Individual compliers cannot identified, but it is possible to describe their observable characteristics (Frandsen et al., 2019). Columns (2)-(3) of Table A10 and Columns (2)-(6) of Table A13 show the shares of defendants satisfying each characteristic for both the whole sample and within the complier population. Stronger effects for defendants in higher income quartiles could be due to a higher concentration of compliers in those subsamples. In Table A10 we can see that is not the case. However, both drug offenses and self-representation seem to be overrepresented in the complier population and, if those are differentially distributed across the income distribution that could be driving the heterogeneity. In Table A13 I can see that even though the shares of defendants charged with property and drug offenses, as well as those that self-represented defendants is very different across subsamples (35% and 15%, 35% and 55%, and 8% compared to 22%), the complier samplings relative to each subsample are quite similar.

6.2 Alternative Identification Approaches

Table A16 in the Appendix explores alternative specifications that account for potential biases from the construction of the leniency measure. I report results using all the judge dummies directly as instruments, using Lasso to pick the most informative dummies (Belloni et al., 2014), as well as the more traditional JIVE measure of Dahl et al., 2014; Dobbie et al., 2018. Across all these different estimations strategies I consistently find a negative relationship between monetary sanctions and financially motivated recidivism.

7 Conclusion

I study the effects of monetary sanctions—fines and court fees—on criminal reoffending in the state of North Carolina. I do this by leveraging the as-if random assignment of cases to district judges that vary in their likelihood to levy monetary sanctions. My findings imply that monetary sanctions, on average, decrease the likelihood of defendants engaging in any future criminal activity (9.1 p.p less likely to reoffend, $p < 0.05$) They strongly deter offenses that entail economic gains (10.3 p.p less likely to reoffend, $p < 0.01$). These effects are driven by defendants living in wealthier neighborhoods (12.3 p.p, $p < 0.01$) but there do not seem to be any criminogenic effects for defendants living in poorer areas (3.7 p.p, SE=0.206).

I expand my study of heterogeneity using recent machine learning methods to identify heterogeneous treatment effects. CDDF's method identifies approximately 60 percent of defendants as susceptible to monetary sanctions. Key determinants of reduced recidivism are consistent with my theoretical framework, namely prior involvement with the criminal justice system (susceptible defendants are 11 p.p less likely to have had a prior involvement with the criminal justice system, $p < 0.01$) and neighborhood income (susceptible defendants live in neighborhoods where the median income is \$8,800 higher, $p < 0.01$). Neighborhood income seems to be particularly relevant for the effectiveness of monetary sanctions: YFSBW's approach identifies an allocation of fines and court fees that prioritizes defendants' with higher incomes as particularly effective (18% additional decrease in reoffending—relative to the estimated effects of a 10 p.p. decrease—, $p < 0.01$).

Again, I do not see evidence of reoffending for poorer defendants, in the sense of them resorting to financially motivated crime in order to supplement their income. However, I do find evidence of monetary sanctions putting them in severe financial distress. For defendants living in poorer neighborhoods there is suggestive evidence of a higher likelihood of them being involved in crimes associated with poverty, such as panhandling, sleeping in public spaces, or driving with a revoked license due to unpaid fines or court fees (12.9 p.p more likely to reoffend, SE=0.115).

Fines and court fees, then, have strong deterrent effects, and are considerably less costly than imprisonment, suggesting caution with recent policies that look to ban monetary sanctions, but also highlighting the potential of monetary sanctions as an alternative to incarceration (instead of just being a supplementary penalty). Additionally, my results suggest that, as an alternative to banning monetary sanctions altogether, it may be welfare enhancing to sanction defendants according to their financial circumstances. However, several important questions remain for future research. My results do not explicitly compare imprisonment and monetary sanctions. Evidence along this margin would be useful to assess the validity of these claims. Moreover, while I provide some evidence of the efficacy of graduating monetary sanctions, it would be interesting to study this along the intensive margin of sanctions, as this is ultimately the relevant dimension for graduation.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2):231–263.

Abrams, D. S. and Fackler, R. (2018). To plea or not to plea: Evidence from north carolina. Working paper, university of pennsylvania.

Ackerberg, D. A. and Devereux, P. J. (2009). Improved JIVE Estimators for Overidentified Linear Models with and without Heteroskedasticity. *The Review of Economics and Statistics*, 91(2):351–362.

Agan, A. Y., Doleac, J. L., and Harvey, A. (2021). Misdemeanor prosecution. NBER Working Paper 28600.

Ahrens, A., Hansen, C. B., and Schaffer, M. (2019). ivlasso and pdslasso: Programs for post-selection and post-regularization ols or iv estimation and inference.

Alm, J. and Shimshack, J. (2014). Environmental enforcement and compliance: Lessons from pollution, safety, and tax settings. *Foundations and Trends in Microeconomics*, 10(4):209–274.

Andresen, M. E. (2018). Exploring marginal treatment effects: Flexible estimation using stata. *Stata Journal*, 18(1):118–158.

Angrist, J. and Frandsen, B. (2019). Machine labor. NBER Working Paper 26584.

Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67.

Bannon, A., Nagrecha, M., and Diller, R. (2010). *Criminal justice debt: A barrier to reentry.* New York: Brennan Center for Justice.

Bayer, P., Hjalmarsson, R., and Pozen, D. (2009). Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics*, 124(1):105–147.

Bearden v. Georgia, 461 U.S. 660 (1983).

Beckett, K. A., Harris, A. M., and Evans, H. (2008). *The assessment and consequences of legal financial obligations in Washington state.* Washington State Minority and Justice Commission Olympia.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, pages 657–681.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650.

Bhuller, M., Dahl, G. B., Løken, K. V., and Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4):1269–1324.

Blomquist, S. and Dahlberg, M. (1999). Small sample properties of liml and jackknife iv estimators: Experiments with weak instruments. *Journal of Applied Econometrics*, 14(1):69–88.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.

Bureau of Justice Statistics (2004). *Survey of Inmates in State and Federal Correctional Facilities*. Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, Washington, DC.

Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6):2754–81.

Carr-Hill, R. and Stern, N. (1979). *Crime, the Police and Criminal Statistics*. London: Academic Press.

Chao, J. C. and Swanson, N. R. (2004). Estimation and testing using jackknife iv in heteroskedastic regressions with many weak instruments.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2020). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

Colgan, B. A. (2017). Graduating economic sanctions according to ability to pay. *Iowa Law Review*, 103:53–112.

Cook, F. (2014). The burden of criminal justice debt in alabama: 2014 participant self-report survey. *Jefferson Countyś Community Corrections Program.*

Council of Economic Advisers (2015). Fines, fees, and bail: Payments in the criminal justice system that disproportionately impact the poor [issue brief].

Dahl, G. B., Kostøl, A. R., and Mogstad, M. (2014). Family welfare cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752.

Davidson, R. and MacKinnon, J. G. (2006). The case against jive. *Journal of Applied Econometrics*, 21(6):827–833.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., and Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–40.

Dušek, L. and Traxler, C. (2021). Learning from Law Enforcement. *Journal of the European Economic Association*, 20(2):739–777.

Earnhart, D. and Segerson, K. (2012). The influence of financial status on the effectiveness of environmental enforcement. *Journal of Public Economics*, 96(9-10):670–684.

Finlay, K., Gross, M., Luh, E., and Mueller-Smith, M. (2022). The impact of financial sanctions in the us justice system: Regression discontinuity evidence from michigan's driver responsibility program.

Frandsen, B. (2020). Testjfe: Stata module to perform test for instrument validity in the judge fixed effects design.

Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2019). Judging judge fixed effects. *National Bureau of Economic Research Working Paper 25528.*

Giles, T. (2022). The (non)economics of criminal fines and fees. *Unpublished Working Paper.*

Goncalves, F. and Mello, S. (2017). Does the punishment fit the crime? speeding fines and recidivism. *Speeding Fines and Recidivism (October 27, 2017).*

Gordon, M. A. and Glaser, D. (1991). The use and effects of financial penalties in municipal courts. *Criminology*, 29(4):651–676.

Gupta, A., Hansman, C., and Riehl, E. (2022). Community impacts of mass incarceration. *Available at SSRN 4033774*.

Harris, A. (2016). *A pound of flesh: Monetary sanctions as punishment for the poor*. Russell Sage Foundation.

Harris, A., Evans, H., and Beckett, K. (2010). Drawing blood from stones: Legal debt and social inequality in the contemporary united states. *American Journal of Sociology*, 115(6):1753–1799.

Harris, A., Evans, H., and Beckett, K. (2011). Courtesy stigma and monetary sanctions: Toward a socio-cultural theory of punishment. *American Sociological Review*, 76(2):234–264.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics*, 3(2):211–255.

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

Heckman, J. J. and Vytlacil, E. J. (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, J. J. and Vytlacil, E. J. (2007). Chapter 70 econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. volume 6 of *Handbook of Econometrics*, pages 4779–4874. Elsevier.

Hull, P. (2017). Examiner designs and first-stage f statistics: A caution.

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

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

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. *Unpublished Working Paper*.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165.

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

Maciag, M. (2020). Addicted to fines. *Fees, Fines, and the Funding of Public Services: A Curriculum for Reform*.

Makowsky, M. (2019). A proposal to end regressive taxation through law enforcement. *The Hamilton Project*, 6.

Mann, R. E., Vingilis, E. R., Gavin, D., Adlaf, E., and Anglin, L. (1991). Sentence severity and the drinking driver: relationships with traffic safety outcome. *Accident Analysis & Prevention*, 23(6):483–491.

Markham, J. M. (2018). Monetary obligations in north carolina criminal cases. *University of North Carolina School of Government Technical report*, pages 572–606.

McDonald, D., Greene, J., and Worzella, C. (1992). *Day fines in American courts: the Staten Island and Milwaukee experiments*, volume 100. US Department of Justice, Office of Justice Programs, National Institute of Justice.

Moffatt, S. and Poynton, S. (2007). The deterrent effect of higher fines on recidivism: driving offences. *Crime and Justice Bulletin*, (106):1–15.

Montiel Olea, J. L. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Unpublished Working Paper*.

Natapoff, A. (2018). *Punishment without crime : how our massive misdemeanor system traps the innocent and makes America more unequal*. Basic Books, New York.

NBC Montana (2018). State supreme court cans Missoula court fee. Retrieved from: https://nbcmontana.com/news/local/state-supreme-court-cans-missoula-court-fee.

Nichol, G. and Hunt, H. (2018). Criminalizing poverty in north carolina: Fines and fees. *North Carolina Central Law Review*, 41:25.

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

Pager, D., Goldstein, R., Ho, H., and Western, B. (2022). Criminalizing poverty: The consequences of court fees in a randomized experiment. *American Sociological Review*.

Phillips, G. D. A. and Hale, C. (1977). The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review*, 18(1):219–228.

Piquero, A. R. and Jennings, W. G. (2017). Research note: Justice system–imposed financial penalties increase the likelihood of recidivism in a sample of adolescent offenders. *Youth Violence and Juvenile Justice*, 15(3):325–340.

Rose, E. (2018). The effects of job loss on crime: evidence from administrative data. *SSRN Electronic Journal*.

Rose, E. K. and Shem-Tov, Y. (2021). How does incarceration affect reoffending? estimating the dose-response function. *Journal of Political Economy*, 129(12).

Sibilla, N. (2022). New Delaware Law Ends Many Fines And Fees That Criminalize Poverty. Retrieved from: `https://www.forbes.com/sites/nicksibilla/2022/10/04/new-delaware-law-ends-many-fines-and-fees-that-criminalize-poverty/?sh=72fba48a57b1`.

Sledge, M. (2018). New Orleans judges to appeal decisions in fines and fees lawsuits. Retrieved from: `https://www.nola.com/news/courts/article_785725a9-d1f4-5c1f-9864-64cb9657fff3.html`.

Stevenson, M. (2017). Breaking bad: Mechanisms of social influence and the path to criminality in juvenile jails. *Review of Economics and Statistics*, 99(5):824–838.

Stevenson, M. and Mayson, S. (2018). The scale of misdemeanor justice. *Boston University Law Review*, 98:731.

Thadani, T. (2018). Criminal justice fees for 21,000 waived. Retrieved from: `http://sfchronicle.com/politics/article/Criminal-justice-system-fees-for-21-000-waived-13175850.php`.

Turner, S. and Petersilia, J. (1996). Day fines in four us jurisdictions. *Santa Monica, CA: The RAND Corporation*.

Tuttle, C. (2019). Snapping back: Food stamp bans and criminal recidivism. *American Economic Journal: Economic Policy*, 11(2):301–27.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., and Wager, S. (2021). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *arXiv preprint arXiv:2111.07966*.

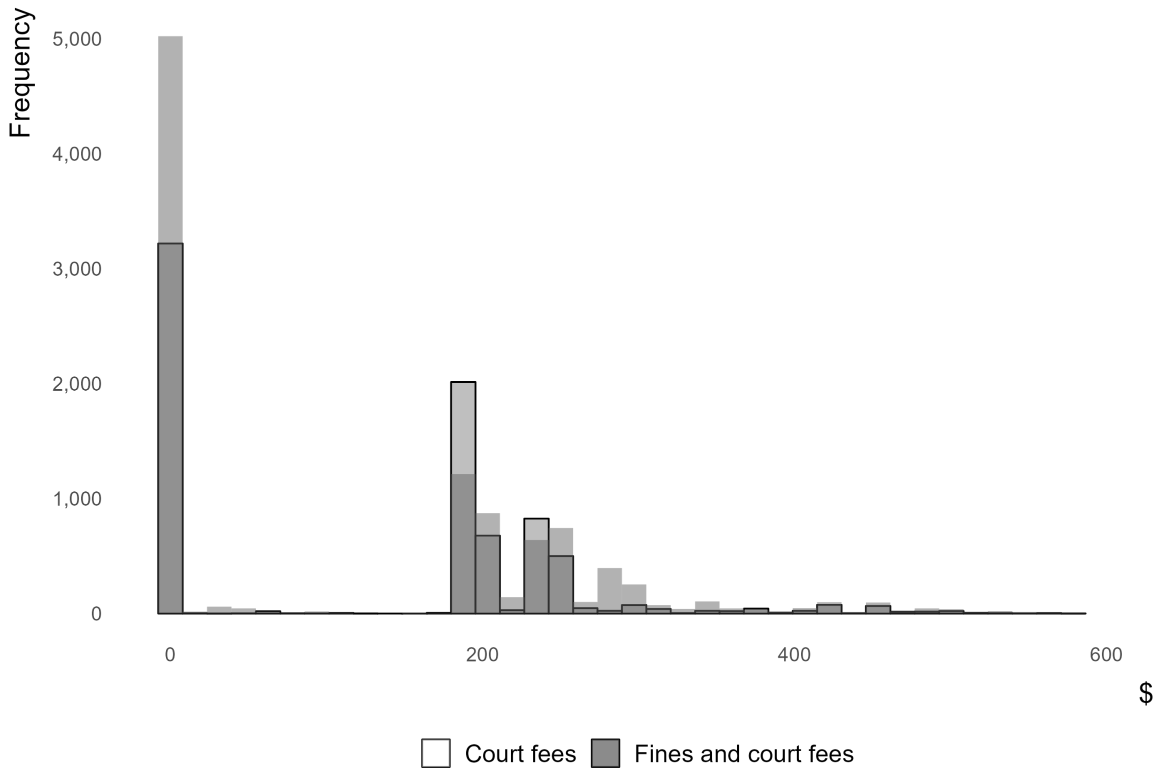
Yu, J. (1994). Punishment celerity and severity: Testing a specific deterrence model on drunk driving recidivism. *Journal of Criminal Justice*, 22(4):355–366.

Zhao, L., Tian, L., Cai, T., Claggett, B., and Wei, L.-J. (2013). Effectively selecting a target population for a future comparative study. *Journal of the American Statistical Association*, 108(502):527–539.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67(2):301–320.

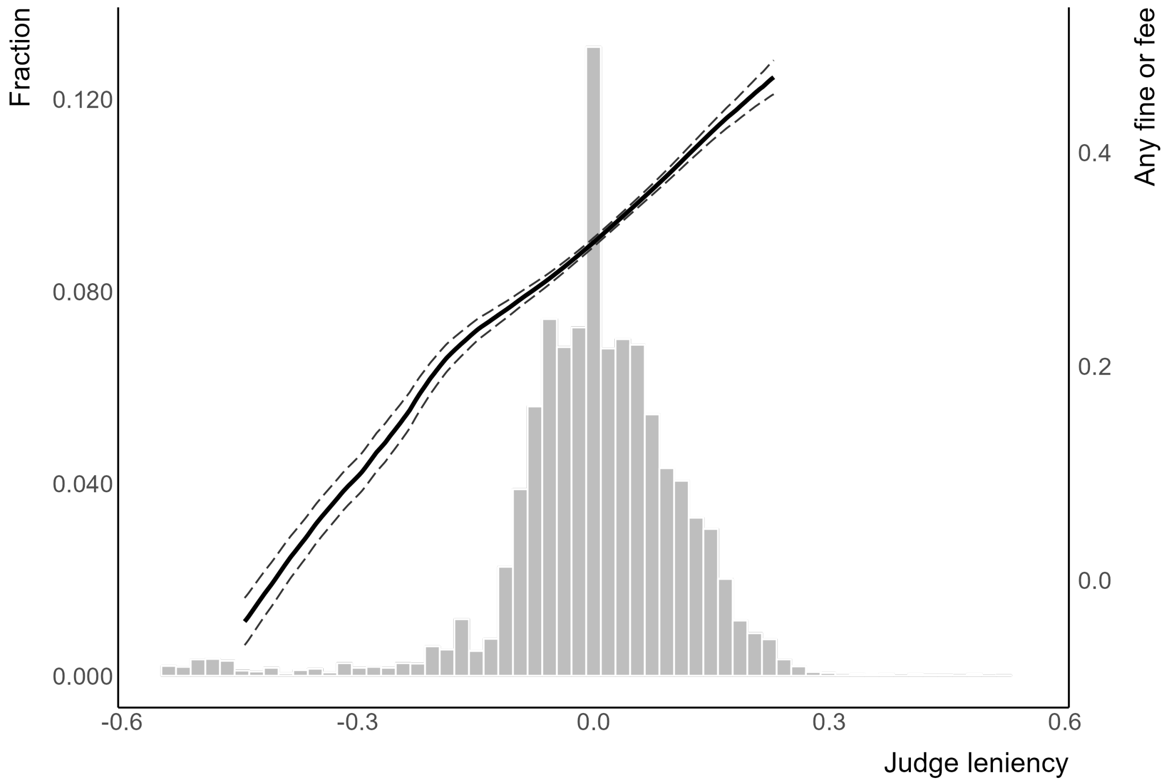
8 Figures and tables

Figure 1: Distribution of Fines and Court Fees



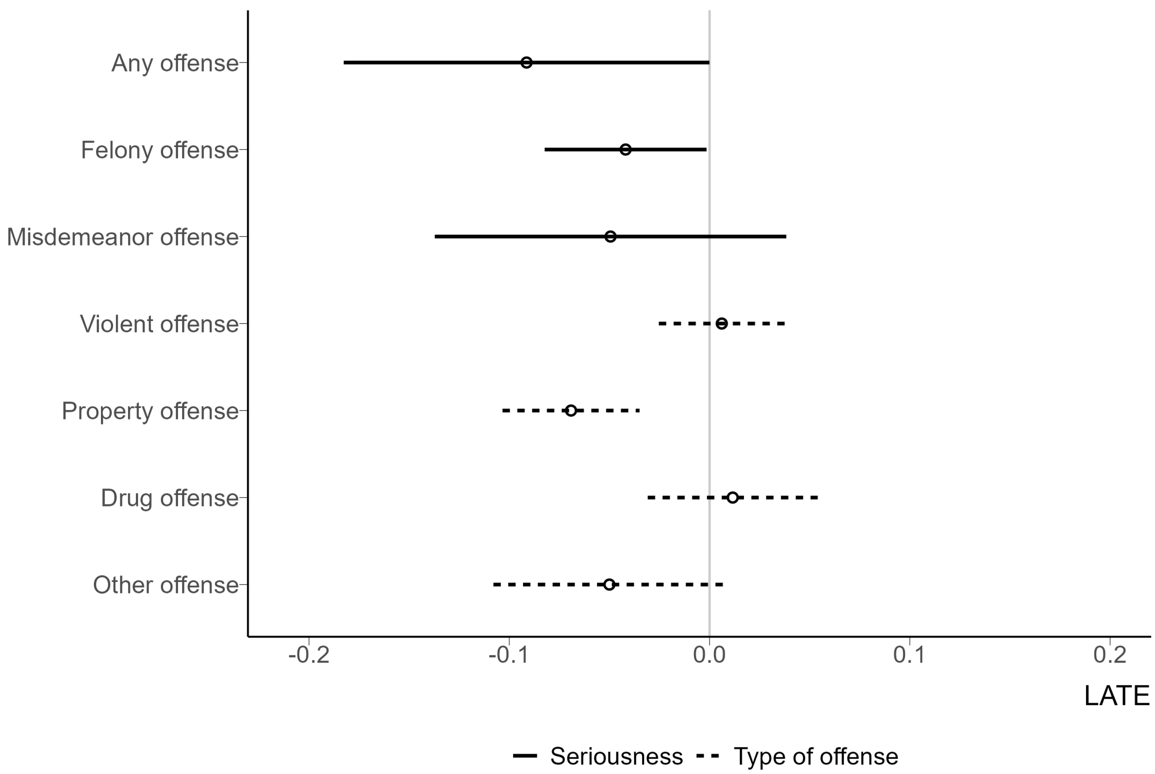
Notes: This figure displays the distribution of fines and court fees in the state of North Carolina for Class 3 misdemeanors for the period 2013–2019.

Figure 2: Distribution of Judge Leniency and First Stage



Notes: This figure shows the distribution of my leave-out mean measure of judge leniency, residualizing out county-by-time fixed effects as described in Section 4. The solid line shows a local linear regression of a dummy for whether a defendant has to pay a fine or court fees on judge leniency, estimated from the 1st to the 99th percentile of judge leniency. Dashed lines show 95% confidence intervals.

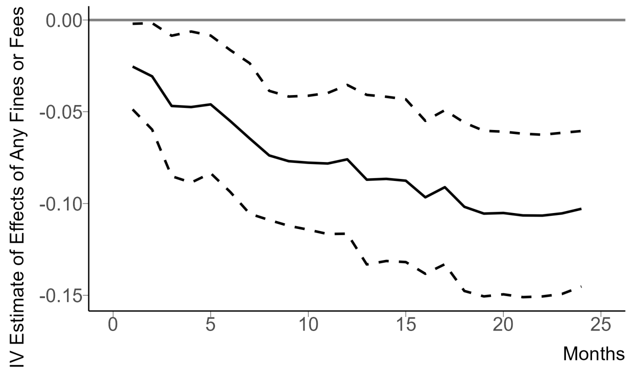
Figure 3: Treatment Effects by Seriousness and Type of Recidivating Offense



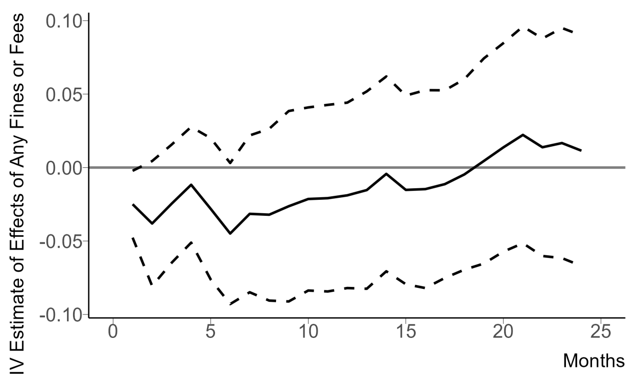
Notes: This figure reports local average treatment effects of being levied any fines or fees on the likelihood of criminal involvement within two years of disposition, for the type of offenses indicated on the y-axis. Estimates are based on the UJIVE estimator of Kolesár (2013), equivalent to column (5) of Table 5. Dashed lines represent 95% confidence intervals clustered at the judge and defendant level.

Figure 4: Treatment Effects Over Time by Financial Reward

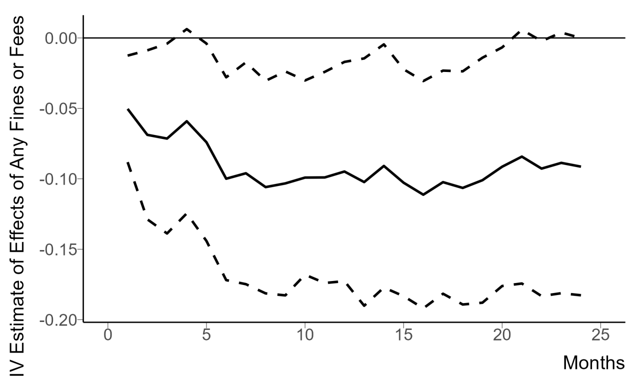
(a) Financially motivated



(b) Non-financially motivated

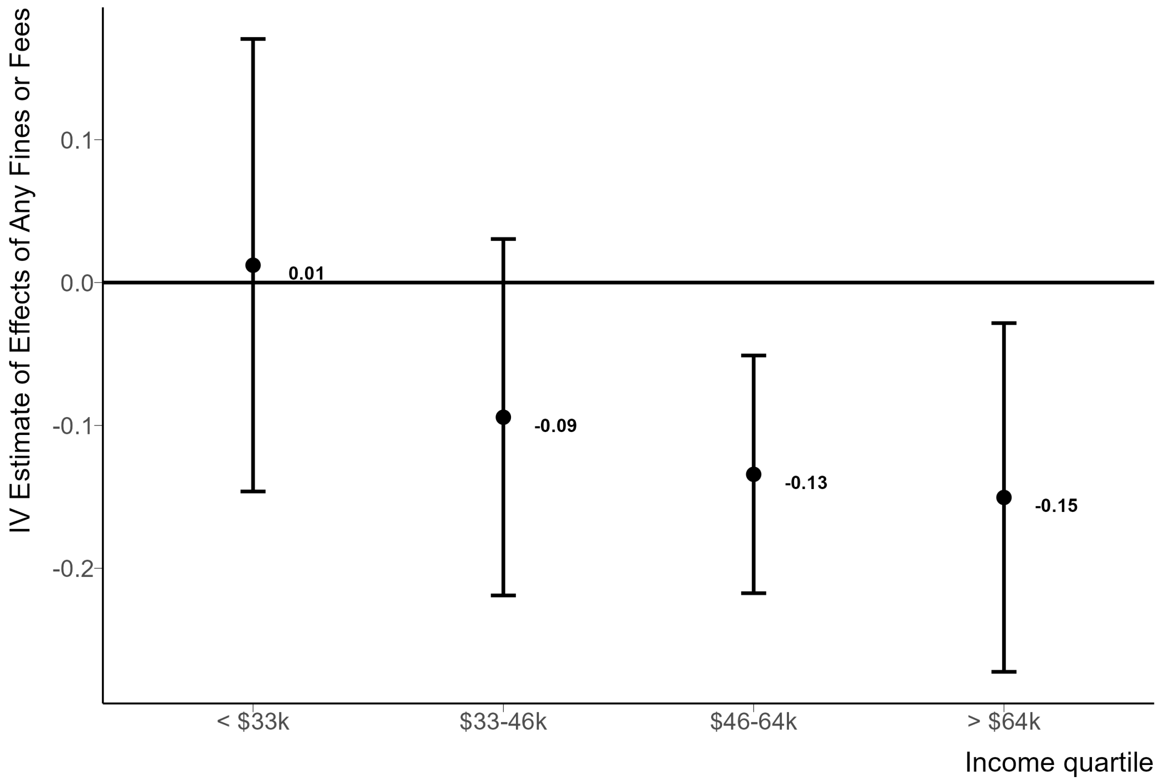


(c) Any offense



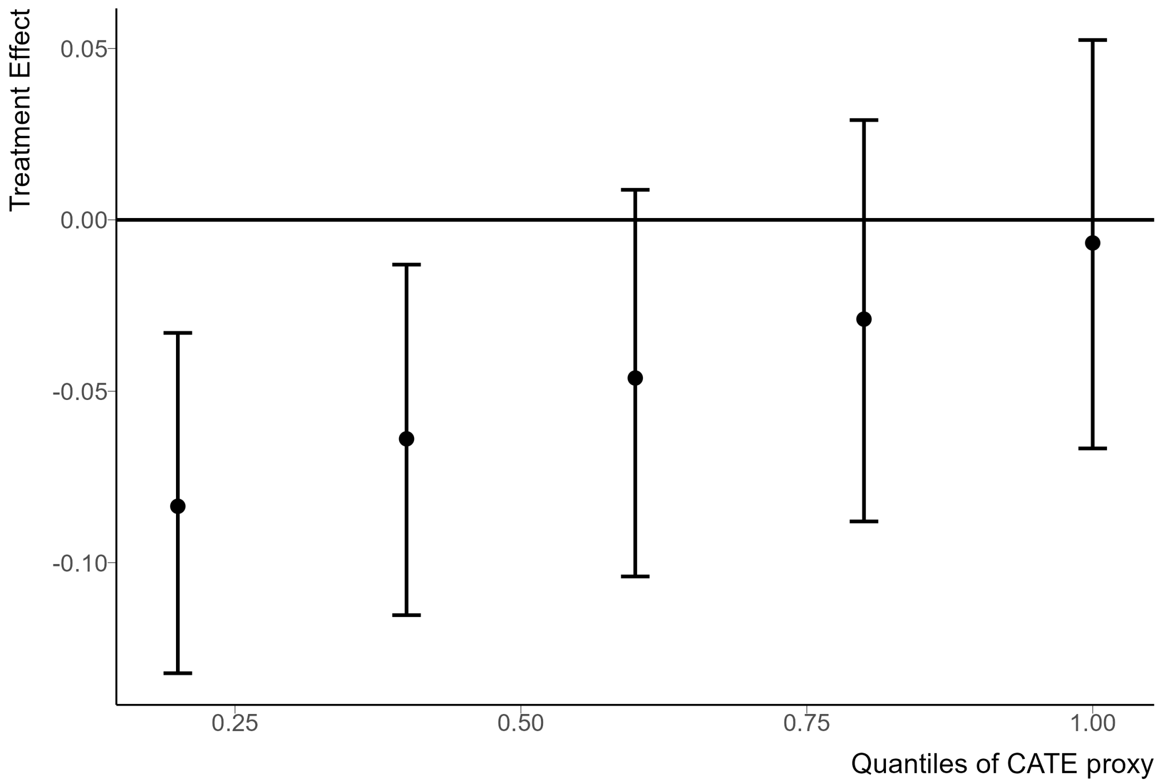
Notes: This figure shows the local average treatment effect of being levied any fines or fees on the likelihood of criminal involvement within a given number of months after disposition. Estimates are based on the UJIVE estimator of Kolesár (2013), equivalent to column (5) in Table 5. Dashed lines represent 95% confidence intervals clustered at the judge and defendant level.

Figure 5: Fines and Fees and Financially Motivated Criminal Involvement Within 2 Years by Census Tract Median Household Income



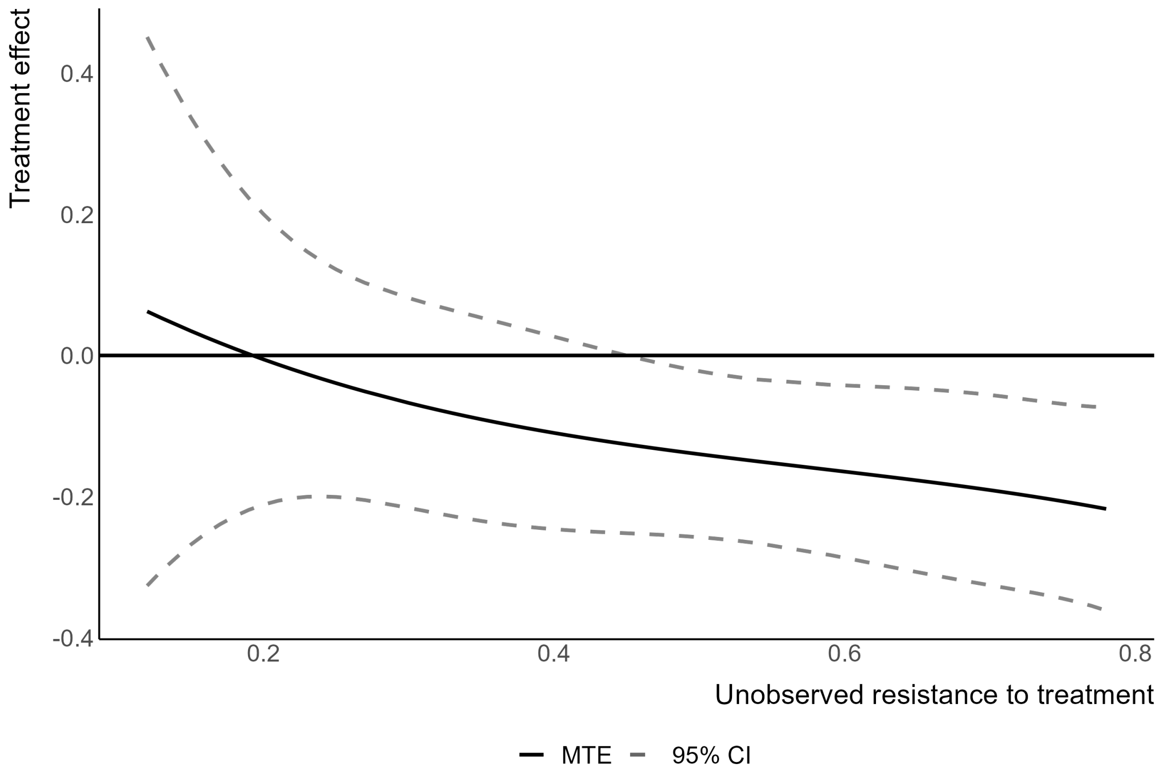
Notes: This figure reports two-stage least squares estimates of being levied any fines or fees on the likelihood of financially motivated criminal involvement within two years after disposition by defendants' income level. Estimates are based on the UJIVE estimator of Kolesár (2013), equivalent of column (5) in Table 5. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. "Income" is proxied by the ACS median household income in the defendant's home address census tract and is in 2019 dollars. Dashed lines represent 95% confidence intervals clustered at the judge and defendant level.

Figure 6: Average Treatment Effects for Quantiles of the Proxy Predictor $S(Z)$



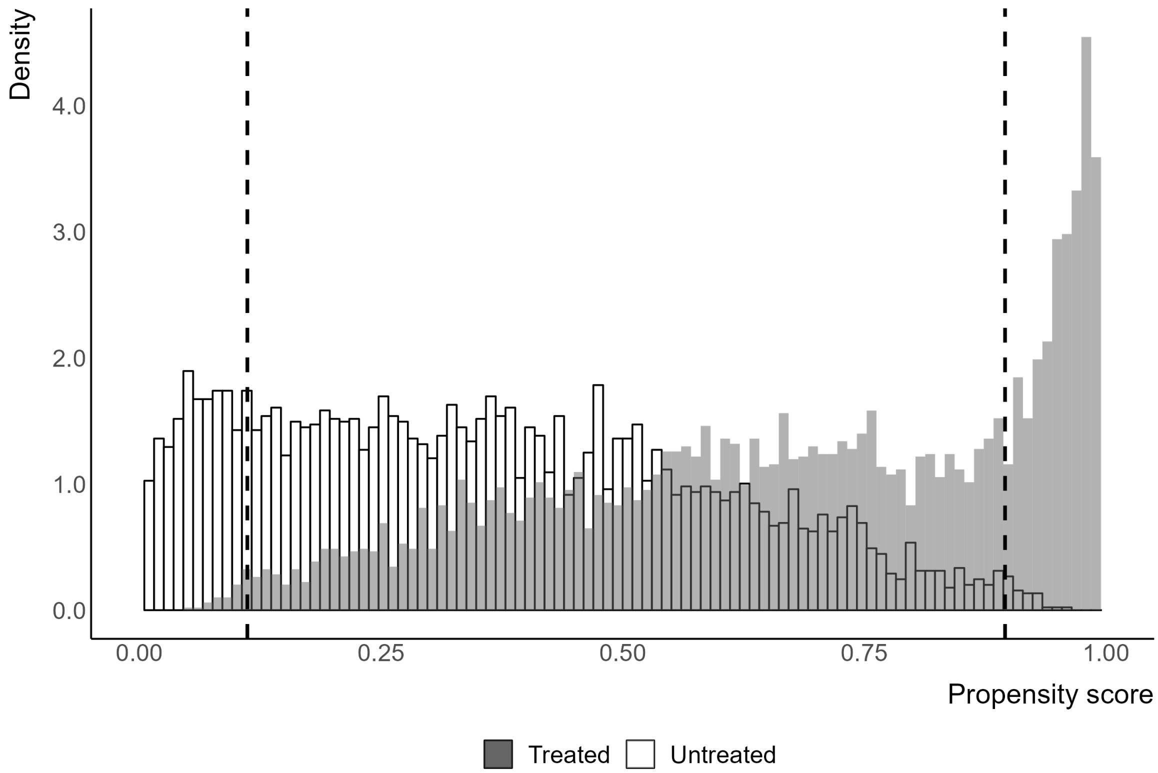
Notes: This figure illustrates heterogeneity in the effects of being levied fines or court fees by plotting estimates of λ_k , $k = 1, \ldots, 5$ from equation 9. Each estimate of λ_k represents the average treatment effect for observations with predicted CATEs lying within the kth quintile of the distribution. The average treatment effect for the entire sample is -0.045 (reported in Table 6). The red bars report 95 percent confidence intervals. Standard errors are tow-way clustered at the individual and judge level.

Figure 7: Marginal Treatment Effects



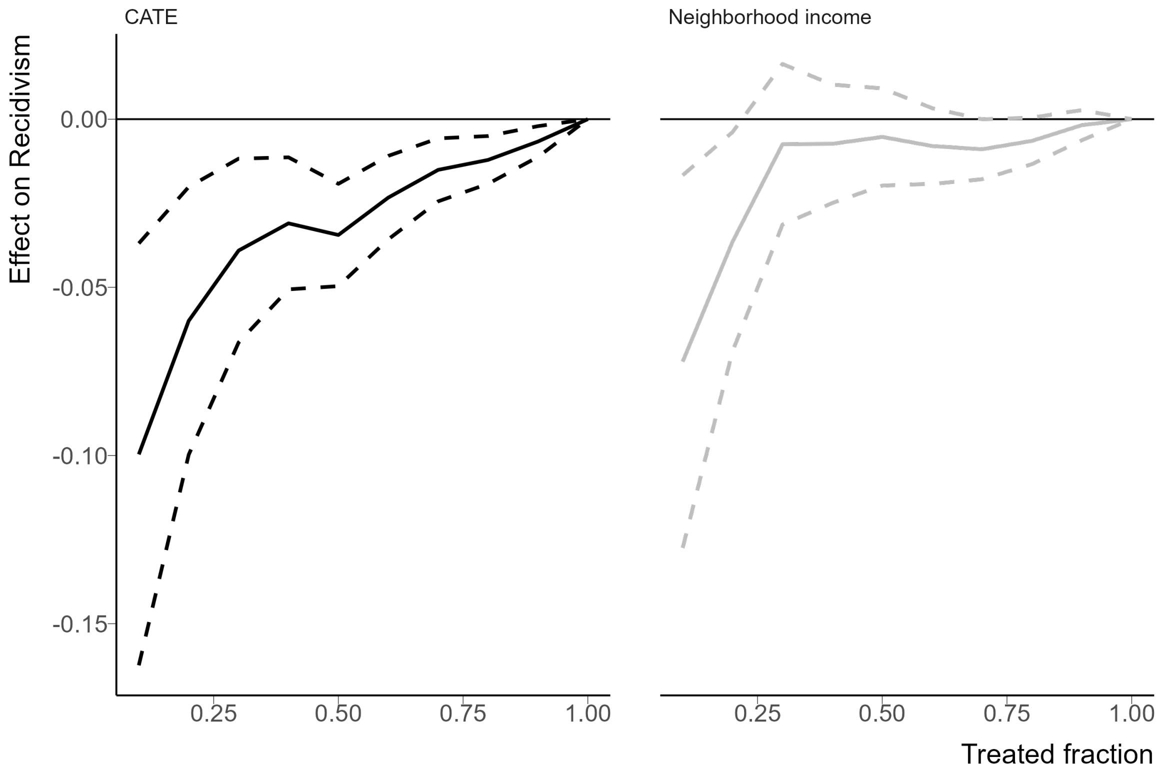
Notes: The MTE estimation is based on a local IV using a cubic polynomial specification in the sample with common support. The x-axis is the predicted probability of receiving a fine, estimated from the assigned district judge after residualizing out covariates and court-by-time fixed effects. Standard errors and resulting 95% confidence intervals are estimated using 100 bootstrap replications. The outcome of interest is the probability of reoffending within two years. All estimations were done using the `mtefe` module in Stata (Andresen, 2018).

Figure 8: Common Support of Propensity Score



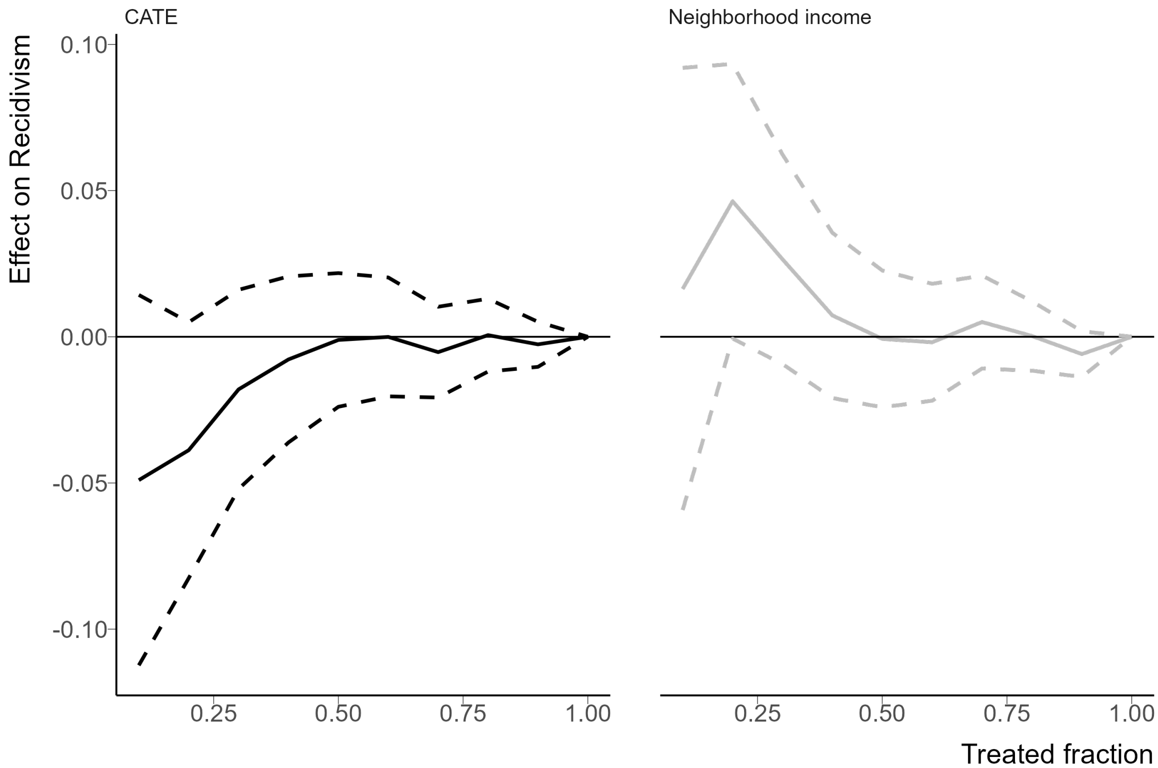
Notes: The dashed lines represent the upper and lower bounds on the common support of the propensity score (based on a 1% trimming) used to estimate the MTEs. Propensity scores are predicted via a logit regression with all case- and defendant-level covariates included, including court-by-time fixed effects. All estimations were done using the `mtefe` module in Stata (Andresen, 2018).

Figure 9: TOC Curves for Financially Motivated Reoffending



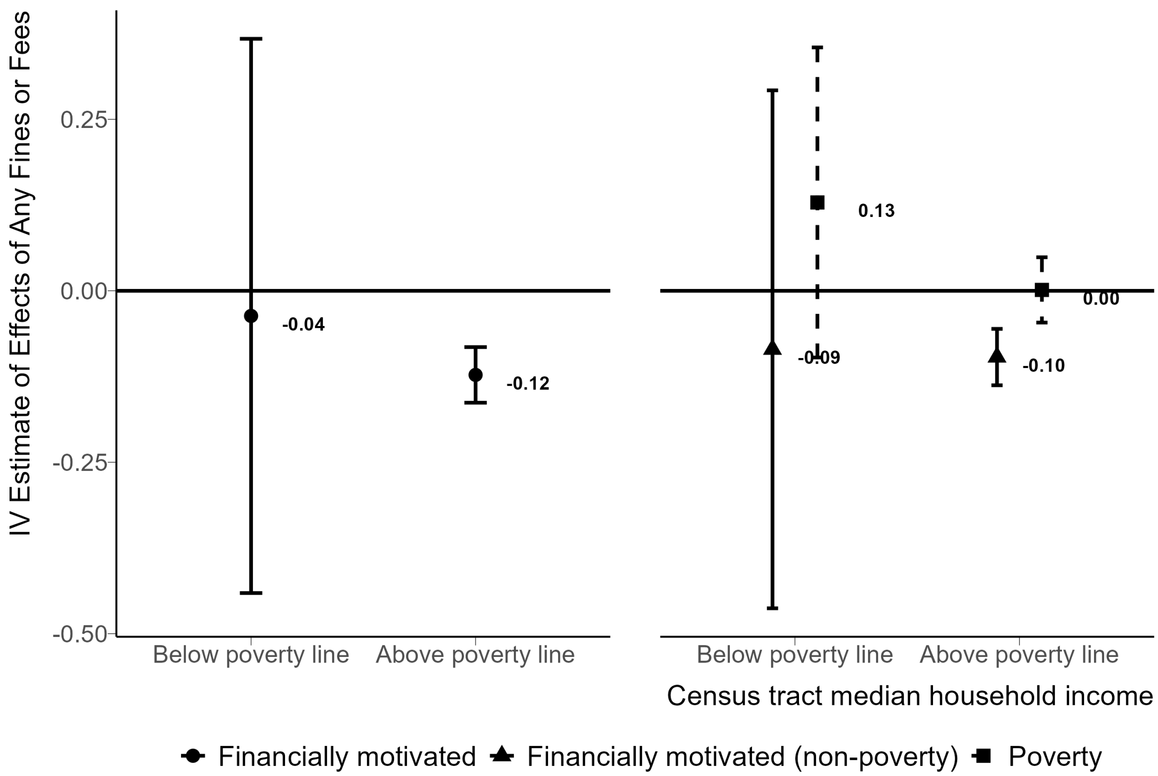
Notes: TOC curves for two prioritization rules (CATE-based and the neighborhood income) for financially-motivated reoffending. These TOC curves correspond to estimates of Equation (10) from the main text and measure differences in treatment effects between individuals in the top q-th fraction of the prioritization rule to the overall average treatment effects.

Figure 10: TOC Curves for Non-Financially Motivated Reoffending



Notes: TOC curves for two prioritization rules (CATE-based and the neighborhood income) for financially-motivated reoffending. These TOC curves correspond to estimates of Equation (10) from the main text and measure differences in treatment effects between individuals in the top q-th fraction of the prioritization rule to the overall average treatment effects.

Figure 11: Treatment Effects by Seriousness and Type of Recidivating Offense



Notes: This figure reports the local average treatment effect of being levied any fines or fees on the likelihood of different types of criminal involvement, within two years after disposition. These are Kolesár (2013)'s UJIVE estimates, equivalent to those of column (5) in Table 5. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. "Poverty" offenses correspond to vagrancy offenses, such as panhandling, begging for money, sleeping in a public place, etc., as well as offenses such as driving with a revoked license due to unpaid traffic tickets. "Income" is proxied by the ACS median household income in the defendant's home address census tract, and is in 2019 dollars. All estimates are shown with 95% confidence intervals clustered at the judge and defendant level.

Table 1: Court Costs Applicable to Analysis Sample Offenses

Item	Amount
Basic Costs	
General Court of Justice Fee	\$147.50
Facilities Fee	\$12.00
Telecom/Data Fee	\$4.00
LEO Retirement Fee	\$7.50
LEO Training Fee	\$2.00
DNA Fee*	\$2.00
Contingent Costs	
Arrest/Process Fee (per service)†	\$5.00
Installment Plan Setup Fee‡	\$20.00
Failure to Appear (FTA) Fee	\$200.00
Failure to Comply (FTC) Fee§	\$50.00
Attorney Fees	
Attorney Fees (per hour)	\$65.00
Attorney Appointment Fee	\$60.00

Sources: Markham (2018), N.C. Administrative Office of the Courts, and N.C. Office of Indigent Defense Services.

* Does not apply to infractions.

† For each arrest or service of criminal process, such as citations or subpoenas.

‡ In addition to any monetary judgment not paid in full.

§ When a defendant fails to pay a fine or court fees within 40 days of the judgment.

Table 2: Case Characteristics and Outcomes

	Mean	Std. Dev.	Observations
Panel A: Defendant Characteristics			
Male	0.73		10,340
Non-Hispanic White	0.36		10,340
African-American	0.57		10,340
Hispanic	0.05		10,340
Other	0.03		10,340
Age	29.95	12.49	10,340
Income ($1,000s)	51.21	24.10	10,340
Panel B: Charge Characteristics			
Number of offenses	1.24	0.43	10,340
Any prior record	0.23		10,340
Previous criminal charges	1.60	0.74	2,400
Type of offense			
Property	0.25		10,340
Drug	0.52		10,340
Other	0.24		10,340
Panel C: Attorney Characteristics			
Court appointed	0.05		10,340
Public defender	0.09		10,340
Privately retained	0.68		10,340
Self-represented	0.18		10,340

	Mean	Std. Dev.	Observations
Panel D: Case Outcomes			
Monetary bail	0.22		10,340
Bail amount ($1,000s)	1.13	5.12	2,248
Convicted	0.69		10,340
Guilty plea	0.67		10,340
Any fine	0.29		10,340
Fine ($)	14.32	33.39	3,20
Any court fees	0.50		10,340
Court fees ($)	113.33	128.02	5,176
Any fine or court fees	0.52		10,340
Fines plus court fees ($)	126.67	136.90	5,332
Unpaid fines or fees	0.38		5,332
Panel E: Recidivism outcomes			
Reoffend within 2 years	0.37		10,340
Type of offense			
Violent	0.10		3,777
Property	0.17		3,777
Drug	0.16		3,777
Other	0.17		3,777

Notes: This table reports descriptive statistics for the analysis sample described in Section 3.2. "Income" is proxied by the ACS median household income in the defendant's home address census tract and is in 2019 dollars.

Table 3: First Stage: Judge Leniency and Fines and Fees

	(1)	(2)
Judge leniency	0.807***	0.720***
	(0.047)	(0.037)
County × Time FE	Yes	Yes
Case/Def. controls	No	Yes
Mean any fines or fees	0.516	
Kleibergen-Paap rk Wald F-Stat	296.81	368.32
Observations	10,340	10,340

Notes: This table reports first-stage results. Judge leniency is estimated using data from other cases assigned to the same district judge following the procedure described in Section 4. Column (1) reports results controlling for the full set of county-by-time fixed effects. Column (2) adds defendant and case controls: defendant race, defendant gender, defendant age, number of offenses, indicators for the type of offense the defendant is charged with, and census block group median income. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses in columns (1) and (2). Robust (Kleibergen-Paap) first stage F reported (which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013) with a single instrument). $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Table 4: Testing for Random Assignment of Cases to Judges

	Any fines or fees	Judge leniency
Age	-0.00108	-0.00040
	(0.00062)	(0.00028)
Male	-0.01528	-0.00328
	(0.01138)	(0.00351)
Minority	0.00540	-0.00497
	(0.01056)	(0.00409)
Number of offenses	0.01294	-0.00556
	(0.01140)	(0.00323)
Property offense	-0.06862∗∗∗	-0.01360∗
	(0.01972)	(0.00663)
Drug offense	0.05895∗	0.01056
	(0.02648)	(0.00580)
Prior record	0.08894∗∗∗	-0.01179∗
	(0.01768)	(0.00536)
Private attorney	0.16486∗∗∗	0.02666∗
	(0.02435)	(0.01234)
Self-represented	0.15046∗∗∗	0.03179
	(0.03293)	(0.01820)
Income below poverty level	-0.03909	0.00851∗
	(0.02502)	(0.00340)
Joint F-test p-value	0.00000	0.27229
Observations	10,340	10,340

Notes: This table reports regressions testing the random assignment of cases to district judges. Judge leniency is estimated following the procedure described in the text. Column (1) reports estimates from an OLS regression of whether the defendant is levied any fines or fees on the variables listed and county-by-time fixed effects. Column (2) reports estimates from an OLS regression of judge leniency on the variables listed and county-by-time fixed effects. A defendant is a minority if their reported race/ethnicity is either Black, Hispanic, or Other. "Income" is proxied by the ACS median household income in the defendant's home address census tract and is in 2019 dollars. Poverty is defined as having a home address in a census block with a median income below the federal poverty income threshold. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. The p-value reported at the bottom of columns (1) and (2) is for an F-test of the joint significance of the variables listed with standard errors two-way clustered at the individual and judge level. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Table 5: Fines and Fees and Criminal Involvement Within 2 years

	OLS			UJIVE	
	(1)	(2)	(3)	(4)	(5)
Panel A: Any criminal involvement within 2 years					
Any fines or fees	–0.022	–0.002	0.000	–0.122***	–0.091**
	(0.014)	(0.013)	(0.013)	(0.041)	(0.047)
Mean of dep. var.	0.365				
Mean of dep. var. for compliers	0.343				
Panel B: Financially motivated criminal involvement within 2 years					
Any fines or fees	–0.054***	–0.035***	–0.033***	–0.145***	–0.103***
	(0.010)	(0.008)	(0.008)	(0.021)	(0.022)
Mean of dep. var.	0.089				
Mean of dep. var. for compliers	0.046				
Panel C: Non-financially motivated criminal involvement within 2 years					
Any fines or fees	0.032***	0.033***	0.033***	0.022	0.011
	(0.009)	(0.010)	(0.010)	(0.035)	(0.040)
Mean of dep. var.	0.277				
Mean of dep. var. for compliers	0.297				
County × Time FE	Yes	Yes	Yes	Yes	Yes
Case/Def. controls	No	Yes	Yes	No	Yes
Complier weights	No	No	Yes	No	No
Observations	10,340	10,340	10,340	10,340	10,340

Notes: This table reports OLS and UJIVE estimates of the impact of fines and fees on the probability of a subsequent criminal involvement within two years. Columns (1) and (4) report results controlling for my full set of county-by-time fixed effects. Columns (2) and (5) add defendant and case controls: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census block group median income. Column (3) reweights Column (2) estimates to match the sample of compliers, as described in Section 5.1. UJIVE estimates instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in Section 4. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. "Income" is proxied by the ACS median household income in the defendant's home address census tract and is in 2019 dollars. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Table 6: Best Linear Predictor of the Conditional Average Treatment Effect

	(1)	(2)	(3)	(4)
β_1 (average treatment effect)	-0.078***	-0.079***	-0.076***	-0.076***
	(0.013)	(0.013)	(0.013)	(0.012)
β_2 (heterogeneity)	0.706***	0.821***	0.588**	0.678**
	(0.273)	(0.292)	(0.258)	(0.273)
Horvitz-Thomson transformation	No	Yes	No	Yes
Trimming threshold	1 %	1 %	5 %	5 %
Observations	9,746	9,746	9,520	9,520

Notes: Columns 1 and 3 present regression estimates of equation (8) from the main text. Columns 2 and 4 present estimates of Equation (12) from the Appendix. The dependent variable is criminal reoffending within 2 years of case disposition. The parameter β_1 measures the average effect of being assigned an stringent judge. Rejecting the null hypothesis that $\beta_2 = 0$ implies that heterogeneity is present and that the proxy predictor, $S(Z)$, captures a component of this heterogeneity. These regressions omit observations with estimated propensity scores less than the trimming threshold or greater than 1 minus the threshold. Standard errors, clustered by defendant, are reported in parentheses.

Table 7: Summary Statistics for Defendants Most and Least Impacted by Fines and Court Fees

	60% Most Affected	20% Least Affected	Difference
Neighborhood income ($)	57,898	48,887	8,825***
			(1,127)
Defendant Characteristics			
Age	27.412	31.575	-4.182***
			(0.590)
Male	0.737	0.728	0.002
			(0.020)
Minority	0.584	0.651	-0.066***
			(0.022)
Any prior record	0.134	0.250	-0.114***
			(0.016)
Charge Characteristics			
Offenses	1.232	1.222	0.004
			(0.019)
Property offense	0.277	0.264	0.020
			(0.020)
Drug offense	0.521	0.448	0.072***
			(0.022)
Other offense	0.193	0.286	-0.089***
			(0.018)

Notes: Column 1 presents means for defendants predicted to have treatment effects in the top 20 percent. Column 2 presents means for defendants with treatment effects in the bottom 20 percent. Column 3 reports the difference between columns 1 and 2. Medians over 100 splits. Standard errors are reported in parentheses.

Table 8: Average Treatment Effects of Receiving a Fine or Court Fee

	Parametric model	Semi-parametric model
ATE	-0.151***	-0.102
	(0.005)	(·)
ATT	-0.107***	-0.124
	(0.001)	(·)
ATUT	-0.071***	-0.047
	(0.001)	(·)

Notes: This Table reports parametric and semiparametric estimates of local average treatment effects (LATE), average treatment effects (ATE), average treatment effects on the untreated (ATUT), and average treatment effects on the treated (ATT), estimated by rescaling the weights on the MTEs for those parameters to integrate over the common support shown in Figure 8. Standard errors are estimated using 100 bootstrap replications.

Table 9: RATE Estimates for Financially and Non-Financially Motivated Reoffending

	AUTOC		Qini	
	Financially motivated	Non-financially motivated	Financially motivated	Non-financially motivated
CATE	-0.036***	-0.014	-0.011***	-0.003
	(0.009)	(0.012)	(0.002)	(0.003)
Neighborhood income	-0.018**	0.007	-0.004*	0.001
	(0.009)	(0.012)	(0.002)	(0.003)

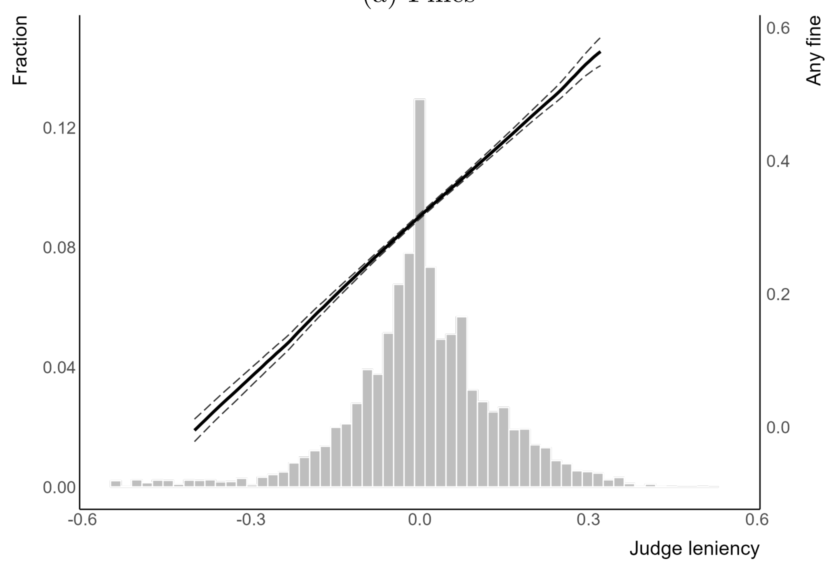
Notes: This table reports RATE metrics' estimates of Equation (11) using both constant (AUTOC) and linear (Qini) weights. Bootstrapped standard errors, 200 replications.

Appendices

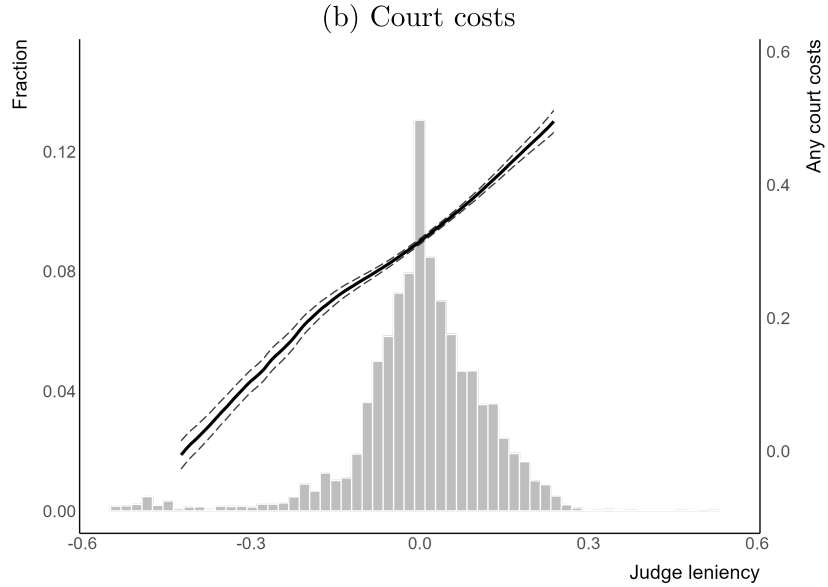
A Additional figures and tables

Figure A1: Distribution of Judge Leniency Levying Fines and Court Fees

(a) Fines

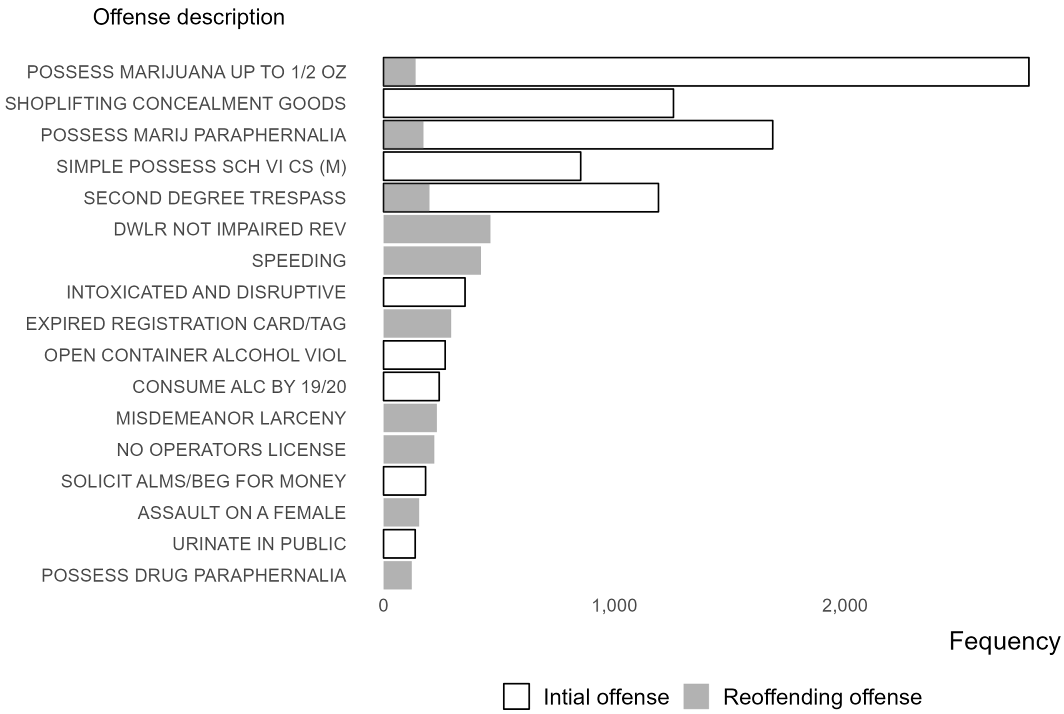


(b) Court costs



Notes: This figure shows the distribution of leave-out mean measures of judge leniency levying both fines and court fees, residualizing out county-by-time fixed effects as described in Section 4. The solid lines show local linear regressions of dummies for whether a defendant has to pay a fine (panel (a)) or court fees (panel (b)) on the respective judge leniency, estimated from the 1st to the 99th percentiles of leniency. Dashed lines show 95% confidence intervals.

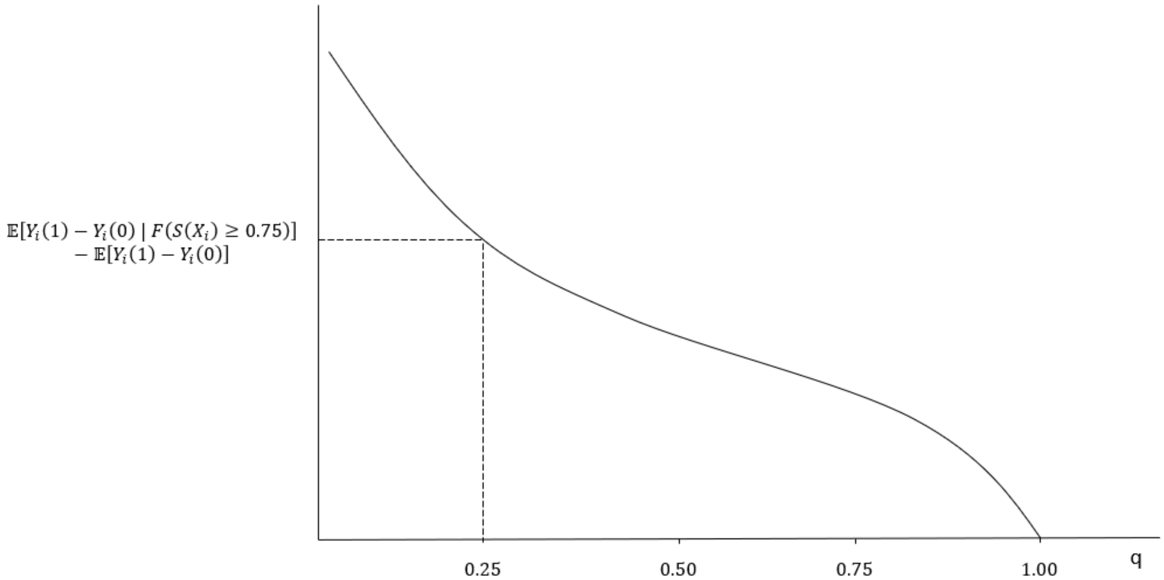
Figure A3: Most Common Initial and Reoffending Offenses



Notes: This figure shows the frequency of the most common initial and reoffending charges defendants face in my sample.

Figure A4: Targeting Operator Characteristic (TOC) Curve

(a) TOC



(b) RATE

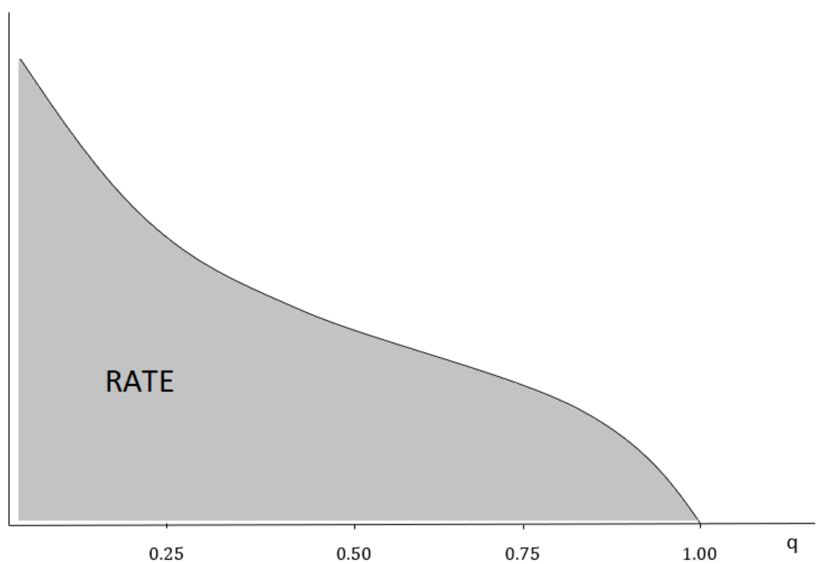


Table A2: Analysis Sample Description

Sample restrictions	Observations
Non motor vehicle misdemeanors	726,696
Class 3 misdemeanors	202,096
Subject only to fines	168,507
Non-waivable court appearance	23,361
No within-County $\times$ Time FE singletons	10,340

Notes: This table reports sample sizes as I restrict the universe of misdemeanor criminal offenses in North Carolina to construct my analysis sample.

Table A3: Case Characteristics and Outcomes by Whether Levied Monetary Sanctions

	All	No fines nor fees	Any fines or fees
Panel A: Defendant Characteristics			
Male	0.73	0.71	0.75
Non-Hispanic White	0.36	0.38	0.34
Minority	0.64	0.62	0.66
Age	29.95	30.39	29.54
	(12.49)	(13.31)	(11.66)
Income ($1,000s)	51.18	47.84	54.30
	(26.31)	(25.35)	(26.80)
Panel B: Charge Characteristics			
Number of offenses	1.24	1.23	1.26
	(0.43)	(0.43)	(0.44)
Any prior record	0.23	0.15	0.31
Property offense	0.25	0.31	0.19
Drug offense	0.52	0.45	0.58
Other offense	0.24	0.24	0.23
Panel C: Attorney Characteristics			
Court appointed	0.05	0.05	0.05
Public defender	0.09	0.13	0.06
Privately retained	0.68	0.69	0.66
Self-represented	0.18	0.14	0.23
Panel D: Case Outcomes			
Any fines or fees	0.52	0.00	1.00
Monetary bail	0.22	0.28	0.16
Convicted	0.69	0.42	0.94
Guilty plea	0.67	0.41	0.92
Monetary bail	0.22	0.28	0.16
Observations	10,340	5,008	5,332

Notes: This table reports descriptive statistics for the analysis sample described in Section 3.2. A defendant is a minority if their reported race/ethnicity is either Black, Hispanic, or Other. "Income" is proxied by the ACS median household income in the defendant's home address census tract and is in 2019 dollars.

Table A4: Relationship between judges, fines and court fees, and conviction

	OLS			UJIVE	
	(1)	(2)	(3)	(4)	(5)
Panel A: Any criminal involvement within 2 years					
Any fines or fees	–0.085***	–0.110***	–0.116***	–0.006	–0.057
	(0.028)	(0.030)	(0.029)	(0.080)	(0.092)
Mean of dep. var.	0.398				
Mean of dep. var. for compliers	0.505				
Panel B: Financially motivated criminal involvement within 2 years					
Any fines or fees	–0.122***	–0.123***	–0.122***	–0.148***	–0.146**
	(0.023)	(0.022)	(0.022)	(0.053)	(0.059)
Mean of dep. var.	0.107				
Mean of dep. var. for compliers	0.051				
Panel C: Non-financially motivated criminal involvement within 2 years					
Any fines or fees	0.037	0.013	0.007	0.142	0.089
	(0.025)	(0.026)	(0.025)	(0.099)	(0.112)
Mean of dep. var.	0.291				
Mean of dep. var. for compliers	0.454				
County × Time FE	Yes	Yes	Yes	Yes	
Case/Def. controls	No	Yes	Yes	No	Yes
Complier weights	No	No	Yes	No	No
Observations	1,732	1,732	1,732	1,732	1,732

Notes: This table reports OLS and UJIVE estimates of the impact of fines and fees on the probability of a subsequent criminal involvement within two years, for those defendants charged with drug or property offenses who have had a prior involvement with the criminal justice system. Columns (1) and (4) report results controlling for my full set of county-by-time fixed effects. Columns (2) and (5) add defendant and case controls: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census block group median income. Column (3) reweights Column (2) estimates to match the sample of compliers, as described in Section 5.1. UJIVE estimates instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in Section 4. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Table A5: Relationship between judges, fines and court fees, and conviction

	OLS		UJIVE	
	(1)	(2)	(3)	(4)
Panel A: Any criminal involvement within 2 years				
Any fines or fees	–0.079***	–0.063***	–0.086**	–0.055
	(0.014)	(0.015)	(0.037)	(0.047)
Mean of dep. var.	0.365			
Panel B: Financially motivated criminal involvement within 2 years				
Any fines or fees	–0.102***	–0.074***	–0.162***	–0.131***
	(0.011)	(0.010)	(0.020)	(0.027)
Mean of dep. var.	0.089			
Panel C: Non-financially motivated criminal involvement within 2 years				
Any fines or fees	0.023**	0.011	0.076**	0.076*
	(0.011)	(0.012)	(0.036)	(0.046)
Mean of dep. var.	0.277			
County × Time FE	Yes	Yes	Yes	Yes
Case/Def. controls	No	Yes	No	Yes
Observations	10,340	10,340	10,340	10,340

Notes: This table reports OLS and UJIVE estimates of the impact of fines and fees on the probability of a subsequent criminal involvement within two years. Columns (1) and (3) report results controlling for my full set of county-by-time fixed effects. Columns (2) and (4) add defendant and case controls: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census block group median income. UJIVE estimates instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in Section 4. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Table A6: Frandsen et al. (2019) Test of Joint Null of Exclusion and Monotonicity

	5 knots	10 knots	15 knots	20 knots
	100.09	98.55	96.51	95.06
	(94)	(89)	(84)	(79)
$\omega = 1.0$	[0.31]	[0.23]	[0.17]	[0.11]
$\omega = 0.8$	[0.39]	[0.29]	[0.21]	[0.13]
$\omega = 0.5$	[0.63]	[0.46]	[0.33]	[0.21]
$\omega = 0.3$	[1.00]	[0.76]	[0.55]	[0.35]

Notes: This table presents results from the test proposed in Frandsen et al. (2019) for the joint null hypothesis that the monotonicity and exclusion restrictions hold. The rows in the first panel report test statistics and degrees of freedom from the fit component of the test. The rows in the second panel report p-values associated with different weighting schemes between the fit and slope components of the test. Each column displays results using a different number of knots in the spline function. A failure to reject the null implies that I cannot reject the hypothesis that the monotonicity and exclusion restrictions jointly hold. This test was implemented in Stata via the package `testjfe` (Frandsen, 2020).

Table A7: First Stage Results by Case and Attorney Characteristics

	Offense type				Attorney type		
	Property	Drug	Other	Financial	Private	Public	Self
Judge leniency	1.050***	0.504***	0.945***	1.007***	0.590***	0.888***	0.918***
	(0.061)	(0.056)	(0.080)	(0.065)	(0.077)	(0.115)	(0.065)
Mean of dep. var.	0.248	0.516	0.236	0.285	0.675	0.141	0.183
Mean of dep. var. for compliers	0.139	0.586	0.275	0.169	0.556	0.050	0.393
County $\times$ Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,418	5,241	2,284	2,810	6,910	1,291	1,785

Notes: This table reports first stage results by offense and attorney type. Judge leniency is estimated using all cases assigned to a district judge in the same year following the procedure described in Section 4. All specifications control for court-by-time fixed effects. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.10$.

Table A8: First Stage Results by Defendant Characteristics

	Prior record		Defendant race		Defendant HH income quintile				
	Any prior	No prior	White	Minority	< \$30k	\$30–41k	\$41–52k	\$52–67k	> \$67k
Judge leniency	0.917***	0.751***	0.799***	0.854***	1.091***	0.932***	0.918***	0.707***	0.657***
	(0.079)	(0.076)	(0.083)	(0.032)	(0.106)	(0.067)	(0.071)	(0.074)	(0.075)
Mean of dep. var.	0.232	0.768	0.359	0.641	0.195	0.195	0.195	0.195	0.220
Mean of dep. var. for compliers	0.193	0.807	0.328	0.672	0.141	0.149	0.187	0.209	0.314
County × Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,253	7,883	3,563	6,504	1,856	1,811	1,842	1,824	2,123

Notes: This table reports first stage results by demographic group and prior criminal involvement. Judge leniency is estimated using all cases assigned to a district judge in the same year following the procedure described in Section 4. All specifications control for court-by-time fixed effects. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$,$**p < 0.05$,$*p < 0.10$.

Table A9: Sample Share by Compliance Type

	Linear model			Local linear model		
	1%	1.5%	2%	1%	1.5%	2%
Compliers	0.60	0.57	0.53	0.54	0.52	0.51
Always takers	0.10	0.12	0.15	0.21	0.21	0.21
Never takers	0.30	0.31	0.32	0.25	0.27	0.28

Notes: This table estimates the shares of the sample that are compliers, always-takers, and never-takers. The fraction of always-takers, π_a, is estimated by the share of the defendants who are fined by the most lenient district judge; the fraction of never-takers, π_n, by the share not fined by the least lenient district judge; and compliers as 1 - π_a - π_n. Most lenient district judges are defined by being at the 1st, 2nd, or 3rd percentile of the residualized district judge leniency distribution, and least lenient are defined as being at the 99, 98, or 97th percentile. The first three columns use a linear specification of the first stage, as in Equation (3); the latter three use a local linear specification.

Table A10: Characteristics of Marginal Defendants

	$\Pr[X = x]$	$\Pr[X = x \mid D(1) > D(0)]$	Ratio
Offenses = 1	0.76	0.81	1.07
Offenses > 1	0.24	0.19	0.78
Any prior record	0.23	0.19	0.83
Property offense	0.25	0.14	0.56
Drug offense	0.52	0.59	1.14
Other offense	0.24	0.28	1.16
Private attorney	0.68	0.56	0.82
Indigent defense	0.14	0.05	0.35
Self-represented	0.18	0.39	2.15
Age $\leq$ 24	0.42	0.36	0.86
Age 25–34	0.25	0.28	1.13
Age 35–44	0.12	0.12	0.98
Age 45–54	0.10	0.10	1.07
Age $\geq$ 55	0.06	0.06	1.02
< \$33k	0.22	0.12	0.56
\$33–46k	0.28	0.27	0.95
\$46–64k	0.27	0.27	0.98
> \$64k	0.22	0.34	1.52
Male	0.73	0.73	1.01
Minority	0.64	0.67	1.05
White	0.36	0.33	0.91

Notes: This table describes the observable characteristics of the complier sample, relative to the full sample. Column (1) shows the probability that an individual has a given characteristic in the full analysis sample. Column (2) shows the probability that someone in the complier group has that characteristic. Column (3) shows the ratio of the two (Column (2) divided by Column (1)). The estimates in Column (2) are constructed by calculating the shares of compliers within these various subsamples. The complier share calculations here rely on a linear first-stage estimation and a 1% cut-off to define district judge leniency.

Table A11: Fines and Fees and Criminal Involvement Within 2 Years by Census Tract Median Household Income

	Household income quartile			
	< \$33k	\$33–46k	\$46–64k	> \$64k
Panel A: Financially motivated criminal involvement within 2 years				
Any fines or fees	0.012	–0.094	–0.134***	–0.150**
	(0.081)	(0.064)	(0.042)	(0.062)
Mean of dep. var.	0.113	0.074	0.075	0.067
Mean of dep. var. for compliers	0.090	–0.041	0.047	0.055
Panel B: Non-financially motivated criminal involvement within 2 years				
Any fines or fees	–0.137	0.107	0.076	–0.100
	(0.122)	(0.099)	(0.089)	(0.093)
Mean of dep. var.	0.245	0.251	0.246	0.261
Mean of dep. var. for compliers	0.225	0.301	0.305	0.229
County × Time FE	Yes	Yes	Yes	Yes
Case/Def. controls	Yes	Yes	Yes	Yes
Observations	2,103	2,740	2,607	2,141

Notes: This table reports local average treatment effects of the impact of fines and fees on the probability of a subsequent criminal involvement within two years by census tract median household income. These are UJIVE estimates that instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in Section 4. All specifications control for case and defendant controls, as well as county-by-time fixed effects. Case and defendant controls are: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census tract median household income. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$,$**p < 0.05$,$*p < 0.10$.

Table A12: Fines and Fees and Criminal Involvement Within 2 Years for Defendants Above and Below the Federal Poverty Line

	Household income	
	< \$26k	≥ \$26k
Panel A: Financially motivated criminal involvement within 2 years		
Any fines or fees	–0.037	–0.123***
	(0.206)	(0.021)
Mean of dep. var.	0.130	0.076
Mean of dep. var. for compliers	0.111	0.026
Panel B: Non-financially motivated criminal involvement within 2 years		
Any fines or fees	0.114	0.019
	(0.286)	(0.040)
Mean of dep. var.	0.251	0.250
Mean of dep. var. for compliers	0.322	0.265
County × Time FE	Yes	Yes
Case/Def. controls	Yes	Yes
Observations	895	9,256

Notes: This table reports local average treatment effects of the impact of fines and fees on the probability of a subsequent criminal involvement within two years, by whether defendant's census tract median household income lies above or below the federal poverty line in North Carolina. These are UJIVE estimates that instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in Section 4. All specifications control for case and defendant controls, as well as county-by-time fixed effects. Case and defendant controls are: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census tract median household income. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$,$**p < 0.05$,$*p < 0.10$.

Table A13: Characteristics of Marginal Defendants Relative to Overall Sample by Whether Household Income is Below or Above Poverty Line

	$<$ \$26k		$\geq$ \$26k	
	Subsample	Compliers	Subsample	Compliers
Offenses = 1	0.77	0.76	0.76	0.82
Offenses > 1	0.23	0.24	0.24	0.18
Any prior record	0.31	0.22	0.44	0.18
Property offense	0.35	0.24	0.15	0.14
Drug offense	0.34	0.54	0.48	0.59
Other offense	0.30	0.23	0.37	0.27
Private attorney	0.73	0.67	0.57	0.56
Indigent defense	0.19	0.14	0.18	0.04
Self-represented	0.08	0.20	0.25	0.40
Age $\leq$ 24	0.27	0.44	0.35	0.36
Age 25–34	0.26	0.25	0.24	0.28
Age 35–44	0.16	0.12	0.27	0.11
Age 45–54	0.17	0.09	0.03	0.11
Age $\geq$ 55	0.11	0.06	0.11	0.06
Male	0.74	0.73	0.72	0.74
Minority	0.68	0.64	0.87	0.66
White	0.32	0.36	0.13	0.34

Notes: This table describes the observable characteristics of the complier sample relative to the full sample for each different quartile of household income, where the latter is proxied by census tract median household income. Each column shows the ratio of the share of compliers, within the subsample defined by the respective covariate, relative to the share for the overall sample. The complier share calculations here rely on a linear first-stage estimation and a 1% cut-off to define district judge leniency.

Table A14: Fines and Fees and Criminal Involvement by Census Tract Median Household Income

	Household income	
	< \$26k	≥ \$26k
Panel A: Financially motivated criminal involvement within 2 years		
Any fines or fees	–0.085	–0.097***
	(0.193)	(0.021)
Mean of dep. var.	0.116	0.070
Mean of dep. var. for compliers	0.106	0.022
Panel B: Poverty-associated motivated criminal involvement within 2 years		
Any fines or fees	0.129	0.001
	(0.115)	(0.024)
Mean of dep. var.	0.051	0.058
Mean of dep. var. for compliers	0.131	0.068
Panel C: Non-financially motivated criminal involvement within 2 years		
Any fines or fees	0.033	–0.008
	(0.263)	(0.038)
Mean of dep. var.	0.214	0.198
Mean of dep. var. for compliers	0.197	0.201
County × Time FE	Yes	Yes
Case/Def. controls	Yes	Yes
Observations	895	9,256

Notes: This table reports local average treatment effects estimates of the impact of fines and fees on the probability of a defendant's subsequent criminal involvement within two years, by whether defendant's census tract median household income lies above or below the federal poverty line in North Carolina. These are UJIVE estimates that instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in the Section 4. All specifications control for case and defendant controls, as well as county-by-time fixed effects. Case and defendant controls are: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census tract median household income. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. "Poverty" offenses correspond to Vagrancy offenses, such as panhandling, begging for money, sleeping in a public place, etc., as well as offenses such as driving with a revoked license due to unpaid traffic tickets. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$,$**p < 0.05$,$*p < 0.10$.

Table A15: Fines and Fees and Criminal Involvement Within 2 Years by Prior Criminal Record

	Defendant prior criminal record	
	Any prior record	No prior record
Panel A: Financially motivated criminal involvement within 2 years		
Any fines or fees	–0.055	–0.134***
	(0.061)	(0.025)
Mean of dep. var.	0.101	0.075
Mean of dep. var. for compliers	0.027	0.033
Panel B: Non-financially motivated criminal involvement within 2 years		
Any fines or fees	0.094	0.007
	(0.140)	(0.053)
Mean of dep. var.	0.265	0.246
Mean of dep. var. for compliers	0.358	0.247
Observations	2,244	7,881

Notes: This table reports local average treatment effects of the impact of fines and fees on the probability of a subsequent criminal involvement within two years, by whether defendants had any prior involvement with the criminal justice system. These are UJIVE estimates that instrument whether a defendant was levied any fines or fees using a judge leniency measure that is estimated using data from other cases assigned to a judge following the procedure described in Section 4. All specifications control for case and defendant controls, as well as county-by-time fixed effects. Case and defendant controls are: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census tract median household income. "Financially motivated" crimes are: property crimes (excluding property damage crimes such as vandalism), selling/manufacturing/distributing drugs, drug trafficking, fraud, forgery, racketeering, prostitution, counterfeiting, crimes containing "sale", or "sell" in the charge description, and begging/panhandling. Crimes with a non-financial motive are defined as all crimes that are not categorized as financially motivated. Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. $***p < 0.01$,$**p < 0.05$,$*p < 0.10$.

Table A16: Alternative IV Strategies

	(1)	(2)	(3)	(4)	(5)
	UJIVE	JIVE	2SLS	LIML	IV Lasso
Any fines or fees	-0.103***	-0.105***	-0.087***	-0.098***	-0.099***
	(0.022)	(0.021)	(0.018)	(0.022)	(0.018)
District × Time FE	Yes	Yes	Yes	Yes	Yes
Attorney type FE	Yes	Yes	Yes	Yes	Yes
Case/Def. controls	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap rk Wald F-Stat	368.32	578.11	2,527.76	2,527.76	148.35
Any fines or fees	0.516				
Observations	10,340	10,340	10,340	10,340	10,340

Notes: This table reports two-stage least squares estimates of the effects of monetary sanctions on financially motivated recidivism using various estimation strategies for the instrument, as indicated in the column headers. All specifications control for case and defendant controls, as well as county-by-time fixed effects. Case and defendant controls are: defendant race, defendant gender, defendant age, number of offenses, prior criminal involvement, indicators for the type of offense the defendant is arrested for, indicators for the type of defense representation, and census tract median household income. The OLS estimate for this specification can be found in Table 5 Column (2), and is -0.033 (SE=0.008). Robust standard errors two-way clustered at the individual and judge level are reported in parentheses. Column (1) repeats my main 2SLS estimates using Kolesár (2013)'s UJIVE estimator. In Column (2) I present estimates using the JIVE estimator. In Column (3) I use all 104 judge dummy variables directly as instruments in the first stage. Column (3) uses limited information maximum likelihood estimation with all of the dummies as instruments as well. Column (5) uses post-lasso from Belloni et al. (2014) to choose the most informative judge dummy variables; the algorithm chooses 22 of the 104 judge dummies as instruments. $***p < 0.01$,$**p < 0.05$,$*p < 0.10$.

Table A17: Comparison of ML Methods

	Elastic Net	Boosting	SVM	Random Forest
Best BLP (Λ)	0.001	0.000	0.000	0.000
Best GATES ($\bar{\Lambda}$)	0.002	0.002	0.002	0.002

Notes: This table compares different ML methods according to the criteria described in Section C.2.1: (1) the correlation between the ML proxy predictor $S(Z)$ and the (true) CATE $s_0(Z)$ (Λ), and (2) the proportion of the variation of $s_0(Z)$ explained by the average $S(Z)$ across GATES groups ($\bar{\Lambda}$). Each estimate corresponds to the median over 100 50-50 splits.

B Data description and analytical sample

B.1 Selected variables' description

- **Fines**. Fines are financial punishments assessed by a judge upon conviction for any level of offense, typically specified in state statutes as a fixed dollar amount or variable range.

- **Court fees**. State laws allow courts to charge criminal defendants fees to recoup justice system costs. These may include charges for the use of a public defender, the cost of summoning expert witnesses, daily charges for incarceration, etc.

- **Restitution**. Restitution are those costs imposed by the court on the defendant to compensate victims who were either injured and/or suffered loss of or damage to property as a result of the defendant's offense.

- **Legal Financial Obligation (LFO)**. Legal financial obligations, or LFOs, are the fines, court fees, and restitution imposed by the court on top of a criminal sentence upon conviction, i.e., LFOs = fines + court fees + restitution. In my sample I do not observe any restitution, LFOs = fines + court fees.

- **Any LFO**. A binary indicator for a defendant being levied any financial obligation, be it a fine, a court fee, or restitution upon conviction.

- **District court district**. Two court systems operate in North Carolina: the Superior Courts and the District Courts. The latter have jurisdiction over cases involving misdemeanor charges. District courts are divided into 41 districts across the state and sit in the county seat of each county.

- **Judge stringency**. I define a measure of judicial stringency that captures a judge's average tendency to impose fines or court fees among the cases they oversee.

- **Neighborhood income**. To define a neighborhood's SES, I use the B1901300 variable, "Median Household Income in the Past 12 Months" from the 5-year American Community Survey dataset for the years 2015-2019 (available at: https://data.census.gov/cedsci/), which reports the median household income in a defendant's census block group.

  Alternatively, I also follow Norris et al. (2021) and Gupta et al. (2022) and use the B17017001 variable from the 5-year 2019 American Community Survey dataset, which measures the share of census block group residents living below the poverty line. For each census block group, I compute the fraction of households with income below the poverty level in the past 12 months. I then rank all census block groups in North Carolina based on this measure, and use the percentile rank as a proxy for socioeconomic status. The value one represents the neighborhood with the lowest poverty rate, and zero represents the neighborhood with the highest poverty rate.

B.2 Data

My main data source is from the North Carolina's Administrative Office of the Courts (NCAOC). These records cover all court cases in the state with information on the defendants, offenses, and sentencing outcomes. I have put together two different 5-year data extracts provided by North Carolina's Automated Criminal/Infractions System (ACIS). These extracts include all cases in which the date of last update was between January 1, 2013 and June 30, 2018 and January 1, 2014 and December 31, 2018, respectively. My analysis focuses on low-level criminal cases—class 3 misdemeanors—and non-criminal infractions. I also have data on felonies, more serious misdemeanors, and criminal traffic offenses such as DWIs and DWLRs (Driving While License Revoked).

Most of the variables in my analysis come from these ACIS' datasets:

1. CASE-RECORD. This dataset includes an observation for each criminal case and includes characteristics of both defendants and their cases. I observe each defendant's race, gender, and age, as well as their exact address at the date of last update in the system. Case characteristics include the date of origination, the court county and type (district/superior), the court session (whether it's a morning or afternoon session, or night court), and the type of bond and its amount, as well as the type of defense attorney (appointed counsel or public defender if the defendant qualifies for indigent defense, or privately retained—or whether the defendant waives their right to an attorney altogether). Additionally, this dataset also includes information regarding defendants' failure to pay their fines/court fees.

2. OFFNS-RECORD. This dataset includes an observation for each charge of each criminal case, with information on the offense and disposition outcome. These records include the date of arrest, arraignment, and disposition, as well as a separate 4-digit offense code for each of these three events. I observe the defendant's plea, verdict, and type of disposition (be it a judge, a jury, or whether the offense was dismissed by the prosecutor). Importantly for my empirical strategy, the dataset includes the judge's initials if a judge was involved in the disposition. Lastly, this dataset also includes information on the sentencing outcome, including the type of sentence (active,

intermediate, community), minimum/maximum sentence lengths, structured sentencing offense class, defendant's prior criminal record, probation, and fines/court fees.

B.3 Data Cleaning and Analytical sample

In the NCAOC data, the most granular unit of observation is a charge (crolno) associated with a particular case number (crrkey). Many cases contain multiple charges, and it is common for multiple case numbers to refer to the same criminal event. Thus an important step in data cleaning is to combine charges and cases that are associated with the same criminal activity. I do this by converting the court data into a defendant panel that has one observation per defendant and disposition event. A disposition event is any update in the defendant's case, including the addition or dismissal of charges, superseding indictments, and judge rulings. Disposition events are identified in the offense records (OFFNS-RECORD) by the date variable crdddt. I create a defendant panel in three steps. First, I create a unique ID number for individual defendants.[69] This identifier is based on the defendant identification variables that I observe in the court data: full name (crrnam), date of birth (crrdob), driver's license number (crrdln), last four SSN digits (crrssn), and street address (defined by crradd, crrcty, crrdst, and crrzip). I assign the same ID number to observations that match exactly on full name and at least one of the other four identification variables. To allow for typos or variations in the recorded name, I also use a fuzzy match of names in combination with the requirement that observations match on at least two of the other identification variables. Second, I use the defendant ID to collapse the court data to a dataset with one observation per defendant/disposition event. If there are multiple charges or case numbers associated with defendant/disposition pair, I keep the observation with the most serious 4-digit offense code at the time of arrest (croffc). I define the most serious offense as the one that has the highest offense class.[70] If there are multiple offenses with the same class, I break ties using the more common 4-digit offense code as computed in the data. Throughout the paper, when I refer to a defendant's "offense"for a given disposition, I am referring to the most serious offense as defined by these criteria.

[69]Creating a defendant ID also allows me to examine defendant recidivism outcomes.

[70]The North Carolina court system uses a structured sentencing scheme that classifies felony charges into ten classes (A, B1, B2, C, D, E, F, G, H, and I, where A is the most severe one and I, the least)

C Technical Appendix

C.1 Understanding Compliers

In this section I describe the details on how I calculate the shares and characteristics of compliers, always-takers, and never-takers.

I follow Dahl et al. (2014), same as Dobbie et al. (2018), Bhuller et al. (2020), and Agan et al. (2021), and first define compliers as defendants for whom the decision of whether to charge them fines or court fees would have been different had their case been assigned to the most lenient instead of the most strict judge. Thus, their share is given by

$$\pi_c = \Pr\left[Any\ LFO_i = 1 \mid Z_i = \bar{z}\right] - \Pr\left[Any\ LFO_i = 1 \mid Z_i = \underline{z}\right] = \Pr\left[Any\ LFO_i\left(\bar{z}\right) > Any\ LFO_i\left(\underline{z}\right)\right]$$

where $\bar{z}$ is the maximum value of the judge instrument (the most strict judge) and $\underline{z}$ represents the minimum value (the most lenient judge).

Always-takers, then, are those defendants who would always be levied fines or fees, regardless of the judge they are assigned to. Now, because of the monotonicity and independence assumptions, the share of always takers is given by the probability of having to pay fines or fees when assigned to the most lenient judge:

$$\pi_a = \Pr\left[Any\ LFO_i = 1 \mid Z_i = \underline{z}\right] = \Pr\left[Any\ LFO_i\left(\bar{z}\right) = Any\ LFO_i\left(\underline{z}\right) = 1\right]$$

Finally, never-takers are defendants who would never be levied fines or fees. The fraction of never-takers is given by the probability of being levied no fines or fees by the most strict judge:

$$\pi_n = \Pr\left[Any\ LFO_i = 0 \mid Z_i = \bar{z}\right] = \Pr\left[Any\ LFO_i\left(\bar{z}\right) = Any\ LFO_i\left(\underline{z}\right) = 0\right]$$

Now, to calculate these shares I begin by defining the "most lenient" judge as the bottom 1 percentile of judge leniency and the "most strict" judge as the top 1 percentile, and using a linear regression specification of the first stage Equation 4. Column (1) of Table A9 reports estimates of this specification, controlling for court-by-time fixed effects. Under this specification, I can recover π_c as $\hat{\alpha}_1(\bar{z} - \underline{z})$, π_a as $\hat{\alpha}_0 + \hat{\alpha}_1\underline{z}$, and π_n as $1 - \hat{\alpha}_0 - \hat{\alpha}_1\bar{z}$ where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are the estimated first stage coefficients. I find that 48 percent of my sample are compliers, 37 percent are never takers, and 16 percent are always takers. Columns (2) and (3) of Table A9 report estimates using alternative cutoffs to define the most lenient and most strict judges. Additionally, to explore the sensitivity of these shares to the model specification, in Columns (4)–(6) of Table A9 I also report estimates from a local linear specification of Equation 4. My results are robust to both the particular model specification and the cutoff choice.

C.2 Heterogeneous treatment effects

C.2.1 Chernozhukov, Demirer, Duflo, and Fernandez-Val (2020)

There is a large recent literature aimed at characterizing treatment effect heterogeneity, and machine learning methods have proven particularly promising where standard approaches were lacking. However, they fail to attain uniformly valid inference and, furthermore, in high dimensional settings, absent strong assumptions (e.g., Donsker conditions), they may not even produce consistent estimators of the conditional average treatment effect (CATE).

Chernozhukov et al. (2020)—henceforth, CDDF—sidestep these issues by focusing not on estimating the CATE itself, but instead on validating a given estimated CATE function and/or testing for the presence of heterogeneity. Their approach is as follows. Following their notation, let $Y(1)$ and $Y(0)$ be the potential outcomes for treated and untreated units, D denote treatment status, Z be a vector of covariates, and $p(Z)$ be the propensity score. The causal functions are

$$b_0(Z) := \mathbb{E}[Y(0) \mid Z],$$

which CDDF dub the baseline conditional average (BCA), and

$$s_0(Z) := \mathbb{E}[Y(1) \mid Z] - \mathbb{E}[Y(0) \mid Z],$$

the CATE. As mentioned above, ML estimators $B(Z)$ and $S(Z)$ of these functions cannot be regarded as consistent, nor do they allow the construction of valid confidence intervals. CDDF suggest, then, to focus on simpler features of $s_0(Z)$ instead of $s_0(Z)$ itself. Specifically, they study the best linear predictor of the CATE and group average treatment effects, where the groups are induced by the ML proxy predictor $S(Z)$, in order to assess the presence of treatment effect heterogeneity while, at the same time, testing the relevance of $S(Z)$ as a predictor of $s_0(Z)$. Additionally, they also advocate comparing the properties of the subpopulations that are most and least affected (according to the groups just identified by partitioning $S(Z)$). Their method consists broadly of 5 steps.

Step 0

First, you split your sample into an auxiliary and main sets.

Step 1

Then, using only observations in the auxiliary set, you construct the proxy predictors $S(Z)$ and $B(Z)$ of the (true) CATE, $s_0(Z)$, and the BCA, $b_0(Z)$, using any of the methods outlined above (and several different base learners for comparison).

Step 2

After constructing the proxy predictors you estimate the following weighted OLS in the main sample:

$$Y = \alpha' X_1 + \beta_1(D - p(Z)) + \beta_2(D - p(Z))(S(Z) - \bar{S}) + \varepsilon \tag{12}$$

where $\mathbb{E}[w(Z)\varepsilon X = 0]$, $X' = [X_1, D - p(Z), (D - p(Z))(S(Z) - \bar{S})]$, $w(Z) = [p(Z)(1 - p(Z))]^{-1}$, $X_1 = [1, B(Z)]$, and $\bar{S}$ is the average of $S(Z)$ over the estimation sample. CDDF show that $\beta_1 + \beta_2(S(Z) - \bar{S})$ is the best linear predictor (BLP) of the CATE.

Step 3

Next, you estimate the sorted group average treatment effects (GATES), $\mathbb{E}[s_0(Z) \mid G]$, where the groups G are defined as disjoint intervals of the proxy predictor, $S(Z)$, using the following weighted OLS in the main sample:

$$Y = \alpha' X_1 + \sum_{k=1}^{K} \gamma_k (D - p(Z)) \mathbb{1}_{S(Z) \in G_k} + \nu \tag{13}$$

where $\mathbb{E}[w(Z)\nu X = 0]$, $X' = [X_1, \{(D - p(Z))\mathbb{1}_{S(Z) \in G_k}\}_{k=1}^{K}]$, $w(Z) = [p(Z)(1 - p(Z))]^{-1}$, $X_1 = [1, B(Z), S(Z), I_k]$, $I_k = [l_{k-1}, l_k)$, and l_k is the (k/K)-th quantile of $S(Z)$ in the main sample. The coefficients of interest, γ_k, then, correspond to the average treatment effect for each group k.

Step 4

Finally, you would estimate the Classification Analysis (CLAN) parameters, comparing the properties of the subpopulations that are most and least affected (the least and most affected groups, G_1 and G_K, for example) as

$$\delta_1 = \mathbb{E}[g(Y, Z) \mid S(Z) \in I_1] \quad \wedge \quad \delta_K = \mathbb{E}[g(Y, Z) \mid S(Z) \in I_K], \tag{14}$$

where $I_k = [l_{k-1}, l_k)$ and l_k is the (k/K)-th quantile of $S(Z)$ in the main sample, same as before, and $g(Y, Z)$ is a vector of unit characteristics.

Step 5

Finally, you compute performance measures for all the base learners you used to estimate the proxy predictors (in terms of the BLP, $\hat{\Lambda}$, and the GATES, $\hat{\bar{\Lambda}}$):

$$\Lambda := |\hat{\beta}_2|^2 \hat{\text{Var}}(S(Z)) \quad \wedge \quad \bar{\Lambda} := \mathbb{E}\left[\sum_{k=1}^{K} \gamma_k \mathbb{1}_{S(Z)\in I_k}\right]^2,$$

where $\Lambda = \text{Corr}^2(s_0(Z), S(Z))\text{Var}(s_0(Z))$, so the base learner with the largest $\hat{\Lambda}$ is the one with the highest correlation with true score $s_0(Z)$.

At the same time, $\bar{\Lambda} = \sum_{k=1}^{K} \gamma_k^2 \mathbb{P}[S(Z) \in I_k]$, which is the part of the variation of $s_0(Z)$ explained by $\bar{S} = \sum_{k=1}^{K} \gamma_k \mathbb{1}_{S(Z)\in I_k}$. So, the ML proxy $S(Z)$ that maximizes $\bar{\Lambda}$ is the one that has the largest R^2 in regressions of $s_0(Z)$ on $S(Z)$ (without a constant).

Wrap up

These 5 steps are then repeated multiple times, each time using a different split of the data. Following this, you:

1. Choose the best ML methods based on the medians of Λ and $\bar{\Lambda}$ over all the splits. In my case these are Elastic Net and Random Forest, as shown in Table A17.

2. Obtain all estimates as the medians over all sample splits and adjust confidence intervals and p-values accordingly ("variational estimation and inference", VEIN,[71] in the paper).

3. Check for the presence of heterogeneity/assess a given CATE estimator by testing the null $\beta_2 = 0$ when estimating Equation (12) (See Table 6).

4. Use the GATES to assess the degree of treatment effect heterogeneity by testing whether the top X% of individuals ranked by their CATE have a different average CATE than, the bottom (1 - X)% (see Figure 6).

5. Explore the sources of treatment effect heterogeneity detected in BLP and GATES via CLAN (see Table 7).

[71] CDDF account for two sources of uncertainty: estimation uncertainty, the sampling variation conditional on the auxiliary sample, and splitting uncertainty, which comes from the random splitting of the data into main and auxiliary samples. For point and interval estimation they report the median over the different splits of the estimated parameters and conditional confidence sets, respectively. They construct p-values by taking the medians of many random conditional p-values as well. They also adjust the nominal confidence levels and p-values to reflect the splitting uncertainty.

C.2.2 Rank-Weighted Average Treatment Effects

When choosing whom to target with a given treatment, that decision is informed by a myriad considerations, such as the predicted effect of the treatment, resource constrains, fairness, etc.

This motivates the question: is this treatment allocation the best treatment allocation? Yadlowsky et al. (2021)—YFSBW hereafter—develop a set of methods that aim to answer this question, to assess different treatment prioritization rules.

The idea is to chop the population up into groups defined by the treatment prioritization (targeting rule/scoring function) one wishes to evaluate—$S(X_i)$ in their notation—,compare the ATE in these groups to the overall ATE, and then aggregate these differences.

YFSBW's method has 3 main steps then. First, choose/estimate a scoring function $S(X_i)$, which prioritizes individuals for treatment in decreasing order of $S(X_i)$. This function may be the estimated CATEs, the realized allocation, some risk score calculated ex ante, etc.

Second, calculate the Targeting Operator Characteristic (TOC), which compares the groups defined by the scoring function to the overall ATE (See Panel A of Figure A4), i.e.,

$$\mathrm{TOC}(q; S) \coloneqq \mathbb{E}[Y_i(1) - Y_0(0) \mid F_S(S(X_i)) \geq 1 - q] - \mathbb{E}[Y_i(1) - Y_0(0)] \tag{15}$$

where we can see that the groups are defined to be the top q-th fraction of individuals with the largest prioritization score.

And, finally, obtain the Rank-Weighted Average Treatment Effects (RATE). The RATE summarizes the different TOCs by integrating the area below the curve defined by them, as illustrated in Panel B of Figure A4:

$$\theta_\alpha(S) = \int_0^1 \alpha(q)\mathrm{TOC}(q; S)dq \tag{16}$$

Different weight functions define different RATEs. For example, with a constant weight, $\alpha(q) = 1$, the RATE corresponds to the area under the TOC curve, AUTOC(S). And using a linear weight function the RATE is equivalent to the are under the Qini curve, the Qini coefficient (Zhao et al. (2013)):

$$\mathrm{AUTOC}(S) = \int_0^1 \mathrm{TOC}(q; S)dq \qquad \mathrm{Qini}(S) = \int_0^1 q\mathrm{TOC}(q; S)dq$$

It is important to note that different weighting schemes put more importance on different groups. The AUTOC places more weight on areas under the curve where the expected treatment benefit is largest. When non-zero treatment effects are concentrated among a small subset of the population this weighting is powerful for testing against a sharp null (AUTOC $= 0$). The Qini coefficient weights the TOC by the fraction treated, which implies placing as much weight on units with low treatment effects as on units with high treatment effects. When non-zero treatment effects are more diffuse across the entire population, this weighting tends to give greater power when testing against a null effect.

C.3 Monotonicity and Exclusion Restrictions

Frandsen et al. (2019)—FLL, henceforth—state that if conditions for LATE identification are satisfied, outcomes averaged at the judge level will be a continuous function with bounded slope of the judge propensity to treat. More specifically, and following their notation, FLL's Theorem 2 asserts that if judge assignment is random, relevant, and it satisfies the exclusion restriction and a monotonicity assumption, then the conditional expectation of the outcome given judge assignment is a continuous function of the judges' propensity to treat, with bounded derivative: $\mathbb{E}[Y_i \mid J_i = j] = \phi(p(j))$, where $\phi \in \mathrm{Lip}_K([0,1])$.

This result comes from rearranging Theorem 1 in Imbens and Angrist (1994), considering an arbitrary "baseline" judge:

$$\mathbb{E}[Y_i \mid J_i = j] = (p(j) - p(1))\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(j) > D_i(0)] + \mathbb{E}[Y_i \mid J_i = 1] \;\; \forall\, j \tag{17}$$

Noting the existence of a marginal propensity, $\bar{p}_i$, for each individual, such that they will be treated when assigned a judge with $p(j) > \bar{p}_i$, the right-hand side of Equation (17) can be written as

$$\phi(j) = (p(j) - p(1))\mathbb{E}[Y_i(1) - Y_i(0) \mid p(1) < \bar{p}_i \leq p(j)] + \mathbb{E}[Y_i \mid J_i = 1]$$

which depends on j only through $p(j)$. Furthermore, it is linear in $p(j)$, and thus continuous. By monotonicity the average slope of $\phi(j)$ through two points p and p' (where $p' \geq p$) can be written:

$$\phi(p') - \phi(p) = (p' - p)\mathbb{E}[Y_i(1) - Y_i(0) \mid p < \bar{p}_i \leq p']$$

which, noting that $\mathbb{E}[Y_i(1) - Y_i(0) \mid p < \bar{p}_i \leq p']$ is bounded by the length of the support of Y.

FLL's test hinges on two observations that follow from this result: first, average outcomes conditional on judge assignment should fit a continuous function of judge propensities; second, the slope of that continuous function should be bounded in magnitude by the width of the outcome variable's support. FLL then combine the fit and slope components via a weighted Bonferroni procedure to produce a single joint test.

The process then is as follows. Regress the outcome Y_i on a flexible function of the judge propensity, $\phi(p(J_i))$. Then, jointly test fit and slope by (i) regressing the residuals from the previous step, $u_i = Y_i - \phi(p(J_i))$, on judge indicators and testing whether the coefficients are jointly zero; (ii) testing whether the slopes of the function are within the bounds given by the support of Y_i.

FLL use B-splines to represent this "flexible" function of judge propensity, which are essentially piecewise polynomial segments strung together at "knots". For a given polynomial order, then, the number of knots determines the function's flexibility.

C.4 Alternative Estimating Strategies

The main analyses in this paper are done using as an instrument a residualized leave-out mean leniency measure that is estimated from the other cases that a judge has decided on. One could use the full set of judge dummies as instruments, but then estimates may be severely biased (Bekker, 1994; Bound et al., 1995; Hausman et al., 2012). Popular solutions to this problem, JIVE and LIML, are biased as well when the number of covariates is large and with heterogenous treatment effects, respectively (Ackerberg and Devereux, 2009; Kolesár, 2013). And even though UJIVE addresses both of these issues it is still not clear how to assess many-weak-instrument bias with constructed instruments such as JIVE or UJIVE (Hull, 2017; Frandsen et al., 2019; Bhuller et al., 2020).

In this subsection I explore the robustness of my main estimates to alternative estimation strategies. These results are shown in Table A16. Column (1) repeats my main UJIVE results and Columns (2)–(4) report estimates using JIVE and 2SLS and LIML using the full set of judge dummies. Column (5) reports estimates using Belloni et al. (2014)'s post-lasso to reduce the number of overidentifying restrictions as an alternative to handle the potential bias from many (weak) instruments.[72]

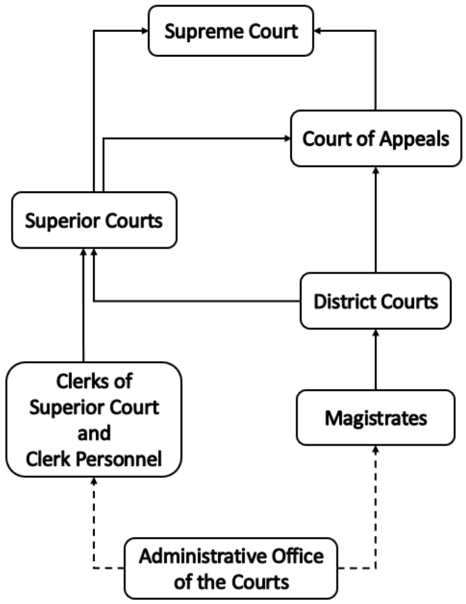
All coefficients are very similar and statistically significant. These estimates imply that within two years following their trial, defendants are between 58% and 95% less likely to reoffend, compared to the sample mean, I cannot reject the null that these coefficients are the same as the UJIVE estimates.[73]

[72] In practice implemented via the user-written package ivlasso in Stata (Ahrens et al., 2019), using the post-lasso results and using the ivlasso defaults with a plug-in penalty (as recommended by Angrist and Frandsen (2019)). The procedure retains 22 out of 104 instruments.

[73] Standard errors reported here for 2SLS or LIML are Montiel Olea and Pflueger (2013)'s robust first-stage F-statistics.

D North Carolina institutional information

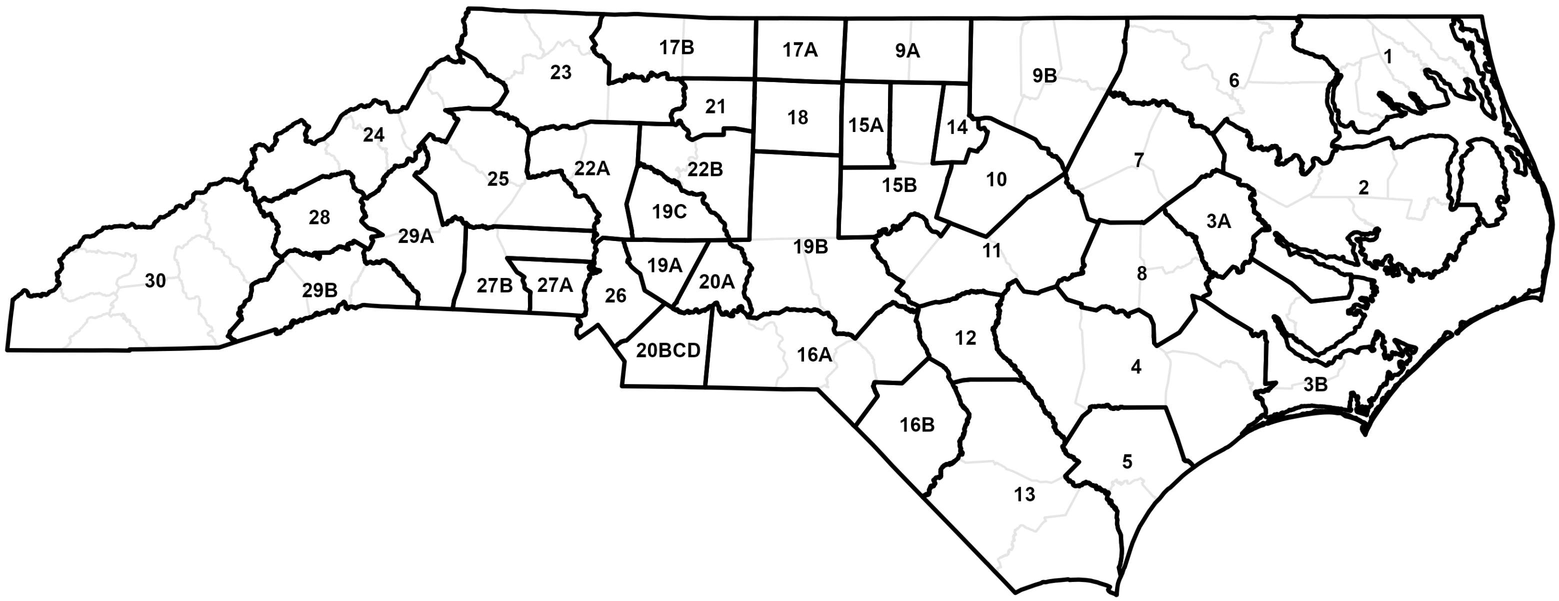
Table C1: North Carolina Judicial System



Judicial Branch Personnel	
Supreme Court justices	7
Court of appeal judges	15
Superior Court judges	107
District Court judges	273
Clerks of superior court	100
District attorneys	43
Public defenders	287

Sources: North Carolina Judicial Branch Statistics—Fiscal Year 2019-20—and North Carolina Office of Indigent Defense Services Workload Assessment, February 2019.

Figure C1: North Carolina District Court Districts



Notes: Effective January 1, 2015–January 1, 2019.

Sources: N.C. General Assembly and UNC Chapel Hill School of Government.

Figure C2: Sentencing Guidelines for Misdemeanors

<table>
<tr><th rowspan="3">CLASS</th><th colspan="4">PRIOR CONVICTION LEVEL</th></tr>
<tr><th>I</th><th colspan="2">II</th><th>III</th></tr>
<tr><th>No Prior Convictions</th><th colspan="2">One to Four Prior Convictions</th><th>Five or More Prior Convictions</th></tr>
<tr><td>A1</td><td>C/I/A
1 - 60 days</td><td colspan="2">C/I/A
1 - 75 days</td><td>C/I/A
1 - 150 days</td></tr>
<tr><td>1</td><td>C
1 - 45 days</td><td colspan="2">C/I/A
1 - 45 days</td><td>C/I/A
1 - 120 days</td></tr>
<tr><td>2</td><td>C
1 - 30 days</td><td colspan="2">C/I
1 - 45 days</td><td>C/I/A
1 - 60 days</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">C
Fine Only*
1 - 10 days</td><td>One to Three Prior Convictions</td><td>Four Prior Convictions</td><td rowspan="2">C/I/A
1 - 20 days</td></tr>
<tr><td>C
Fine Only*
1 - 15 days</td><td>C/I
1 - 15 days</td></tr>
</table>

***Unless otherwise provided for a specific offense, the judgment for a person convicted of a Class 3 misdemeanor who has no more than three prior convictions shall consist only of a fine.**

A – Active Punishment I – Intermediate Punishment C – Community Punishment
Cells with slash allow either disposition at the discretion of the judge

Sources: N.C. General Assembly and UNC Chapel Hill School of Government.

Figure C3: Court Fees Schedule

COURT COSTS AND FEES CHART

The chart below shows court costs in effect as of **December 1, 2019**[1] and applies to all costs assessed or collected on or after that date, except where otherwise noted, and unless subject to the "waiver exception" of G.S. 7A-304(g).

CRIMINAL COURT COSTS G.S. 7A-304, unless otherwise specified			**AMOUNT**
An additional summary chart of criminal costs has been attached to this cost chart as "Appendix - Criminal Costs Summary." The appendix summarizes the basic costs common to all dispositions in a particular trial division. It does **not** include additional cost items that must be assessed depending on individual factors for each case (*e.g.,* FTA fees, supervision fees, jail fees, etc.) or for specific offenses of conviction (*e.g.* improper equipment or impaired driving); those costs are assessed separately. Neither does it apply to offenses for which the relevant statute assesses specific costs or prohibits the imposition of costs.			
DISTRICT COURT (including criminal cases before magistrates)			
General Court of Justice Fee. G.S. 7A-304(a)(4).	General Fund	146.55	147.50
	State Bar Legal Aid Account (LAA)	.95[2]	
Facilities Fee. G.S. 7A-304(a)(2).			12.00
Telecommunications and Data Connectivity Fee. G.S. 7A-304(a)(2a).[3]			4.00
LEO Retirement/Insurance. G.S. 7A-304(a)(3) & (3a).			7.50
LEO Training and Certification Fee. G.S. 7A-304(a)(3b).			2.00
TOTAL			173.00
Chapter 20 Fee. G.S. 7A-304(a)(4a) (for conviction of any Chapter 20 offense).			+10.00[4]
DNA Fee. G.S. 7A-304(a)(9) (criminal offenses, only; does not apply to infractions).			+2.00
Plus $5.00 service fee for each arrest or service of criminal process, including citations and subpoenas. G.S. 7A-304(a)(1).			+5.00
SUPERIOR COURT			
General Court of Justice Fee. G.S. 7A-304(a)(4).	General Fund	153.55	154.50
	State Bar Legal Aid Account (LAA)	.95[5]	
Facilities Fee. G.S. 7A-304(a)(2).			30.00
Telecommunications and Data Connectivity Fee. G.S. 7A-304(a)(2a).			4.00
LEO Retirement/Insurance. G.S. 7A-304(a)(3) & (3a).			7.50
LEO Training and Certification Fee. G.S. 7A-304(a)(3b).			2.00
TOTAL			198.00[6]
Chapter 20 Fee. G.S. 7A-304(a)(4a) (for conviction of any Chapter 20 offense).			+10.00
DNA Fee. G.S. 7A-304(a)(9) (criminal offenses, only; does not apply to infractions).			+2.00
Plus $5.00 service fee for each arrest or service of criminal process, including citations and subpoenas.			+5.00

Figure C4: Court Fees Schedule (continued)

OTHER CRIMINAL FEES	AMOUNT
Appointment of Counsel Fee for Indigent Defendants. G.S. 7A-455.1.	60.00
Certificate of Relief Fee. G.S. 15A-173.2(h).[7]	50.00
Civil Revocation Fee (impaired driving CVRs, only). G.S. 20-16.5(j).	100.00
Community Service Fee. G.S. 143B-708.	250.00
Continuous Alcohol Monitoring (CAM) Fee (offenses prior to Dec. 1, 2012). G.S. 20-179.[8]	Varies[9]
Continuous Alcohol Monitoring (CAM) Fee (parolees, only). G.S. 15A-1374.[10]	Varies
Criminal Record Check Fee. G.S. 7A-308(a)(17).	25.00
Dispute Resolution Fee. G.S. 7A-38.3D and G.S. 7A-38.7.	60.00 per mediation
Expunction Fee, petitions under G.S. 15A-145, 15A-145.1, 15A-145.2, 15A-145.3, 15A-145.4, and 15A-145.7.	175.00
Expunction Fee, petitions under G.S. 15A-145.5.	175.00
Expunction Fee, petitions under G.S. 15A-146.[11]	175.00
Failure to Appear Fee. G.S. 7A-304(a)(6).	200.00
Failure to Comply Fee. G.S. 7A-304(a)(6).	50.00
House Arrest with Electronic Monitoring (EHA) One-Time Fee. G.S. 15A-1343(c2).	90.00
House Arrest with Electronic Monitoring (EHA) Daily Fee. G.S. 15A-1343(c2).	4.48/day
Impaired Driving Fee. G.S. 7A-304(a)(10). **Note:** Applies only to offenses committed on or after December 1, 2011.	100.00
Improper Equipment Fee. G.S. 7A-304(a)(4b).[12]	50.00
Installment Payments Fee. G.S. 7A-304(f).	20.00
Jail Fees (pre-conviction). G.S. 7A-313.	10.00 per 24 hours or fraction thereof
Jail Fees (split sentence served in local facility). G.S. 7A-313 and G.S. 148-29.	40.00 per day
Limited Driving Privilege Fee – Petitions under G.S. 20-20.1. At petition/Application: If Issued: (G.S. 20-20.2).	CVD Costs +100.00
Limited Driving Privilege Fee – Other than under G.S. 20-20.1. If Issued: (G.S. 20-20.2) Note: If there is no underlying conviction in the county, Charge civil filing fees as explained on form AOC-CV-350.	+100.00
Pretrial Release Service Fee (county). G.S. 7A-304(a)(5).[13]	15.00
Satellite-Based Monitoring Fee for Sex Offenders. G.S. 14-208.45.	90.00
State Crime Lab Fee. G.S. 7A-304(a)(7).	600.00
Local Government Lab Fee. G.S. 7A-304(a)(8).	600.00
Private Hospital Lab Fee. G.S. 7A-304(a)(8a).[14]	600.00
State Lab Analyst Expert Witness Fee. G.S. 7A-304(a)(11).[15]	600.00
Local Lab Analyst Expert Witness Fee. G.S. 7A-304(a)(12).[16]	600.00
Private Hospital Analyst Expert Witness Fee. G.S. 7A-304(a)(13).[17]	600.00
State Crime Lab Digital Forensics Fee. G.S. 7A-304(a)(9a).[18]	600.00
Local Lab Digital Forensics Fee. G.S. 7A-304(a)(9b).[19]	600.00
Seat Belt Violations (adult, front seat) and Motorcycle/Moped Helmet Violations. G.S. 20-135.2A and G.S. 20-140.4.	25.50 fine +costs below:
General Court of Justice Fee, G.S. 7A-304(a)(4).	147.50 (Dist.) 154.50 (Sup.)
Telecommunications and Data Connectivity Fee. G.S. 7A-304(a)(2a).	4.00
LEO Training and Certification Fee, G.S. 7A-304(a)(3b).	2.00
TOTAL	179.00 (Dist.) 186.00 (Sup.)
Seat Belt Violations (adult, rear seat). G.S. 20-135.2A(e).	No Costs 10.00 fine only
Supervision Fee. G.S. 15A-1343, G.S. 15A-1368.4, and G.S. 15A-1374.	40.00 per month
Worthless Check Program Fee. G.S. 7A-308(c).[20]	60.00