MISDEMEANOR PROSECUTION*

Amanda Agan† Jennifer L. Doleac‡ Anna Harvey§

August 26, 2022

Communities across the United States are reconsidering the public safety benefits of prosecuting nonviolent misdemeanor offenses, yet there is little empirical evidence to inform policy in this area. In this paper we report the first estimates of the causal effects of misdemeanor prosecution on defendants' subsequent criminal justice involvement. We leverage the as-if random assignment of nonviolent misdemeanor cases to Assistant District Attorneys (ADAs) who decide whether a case should be prosecuted in the Suffolk County District Attorney's Office in Massachusetts. These ADAs vary in the average leniency of their prosecution decisions. We find that, for the marginal defendant, nonprosecution of a nonviolent misdemeanor offense leads to a 53% reduction in the likelihood of a new criminal complaint, and to a 60% reduction in the number of new criminal complaints, over the next two years. These local average treatment effects are largest for defendants without prior criminal records, suggesting that averting criminal record acquisition is an important mechanism driving our findings. We also present evidence that a recent policy change in Suffolk County imposing a presumption of nonprosecution for nonviolent misdemeanor offenses had similar beneficial effects, decreasing the likelihood of subsequent criminal justice involvement.

*We thank former Suffolk County District Attorney Rachael Rollins and the Suffolk County District Attorney's Office for their cooperation. Thanks to Rebecca Regan for excellent research assistance, to Manudeep Bhuller for code and discussions about calculating the complier means, to Martin Andresen for extensive help with the estimation of marginal treatment effects via `mtefe`, and to Mauricio Caceres-Bravo for help with estimation of UJIVE via `manyiv`. We thank Paul Goldsmith-Pinkham, Peter Hull, Michal Kolesár, Emily Leslie, Justin McCrary, Sam Norris, and Roman Rivera for conversations that improved the paper. We appreciate feedback from Jim Greiner, Steve Lehrer, James MacKinnon, Steven Raphael, Jeff Smith, and Megan Stevenson; seminar participants at American University, Bentley University, Boston University, the Federal Reserve Bank of Minneapolis, Georgia State University, LSE-Centre for Economic Performance, Michigan State University, Queen's University, Rutgers University, University of Pittsburgh, University of Toronto, University of Virginia School of Law, and West Virginia University; and conference participants at the 2020 Emory University Conference on Institutions and Law Making, the 2020 APPAM Fall Research Conference, and the 2020 Duke University Empirical Criminal Law Roundtable. We gratefully acknowledge funding from the W.T. Grant Foundation and the Chan Zuckerberg Initiative.

†Department of Economics, Rutgers University, amanda.agan@rutgers.edu

‡Department of Economics, Texas A&M University, jdoleac@tamu.edu

§Department of Politics, New York University, anna.harvey@nyu.edu

1 Introduction

Every year approximately 13 million Americans are charged with misdemeanor offenses, and misdemeanor cases make up over 80 percent of all criminal cases in the United States (Stevenson and Mayson 2018). Many individuals' first contact with the criminal justice system is through misdemeanor charges; in fact, most who enter the criminal justice system will never experience a felony charge.[1] A large proportion of misdemeanor offenses involve neither violence nor firearms, stemming instead from the criminalization of relatively common behaviors such as (depending on the jurisdiction) possession of small quantities of prohibited substances, disorderly conduct, disturbing the peace, trespassing, petty theft, and driving without a valid license/registration/insurance (Stevenson and Mayson 2018; Kohler-Hausmann 2013; Natapoff 2018).

The large volume of misdemeanor cases in the United States raises an important, policy-relevant question about the consequences of misdemeanor prosecution. In most jurisdictions, misdemeanor prosecution implies that a defendant will acquire a criminal record of a misdemeanor charge, even if they are not convicted of that charge. If prosecutors decline to prosecute a defendant's case, the defendant will not acquire a criminal record of a charge. The decision to prosecute a defendant's case may increase "specific deterrence" (Becker 1968) by increasing the punitiveness of misdemeanor defendants' post-arrest experience, thereby decreasing defendants' likelihood of engaging in post-arraignment criminal behavior. But the decision to prosecute may instead increase the likelihood of post-arraignment criminal behavior: because criminal records can carry collateral consequences in multiple spheres (Pager 2008; Uggen et al. 2014; Agan and Starr 2018; Leasure 2019; Natapoff 2018), prosecuted defendants who have acquired criminal records of misdemeanor charges at arraignment may

[1] For example, in Pennsylvania between 2008-2018, 74% of cases for first-time defendants had no felony charges (69% overall). In Bexar County, Texas between 1980-2018, 86% of cases for first-time defendants had no felony charges (79% overall). 62% of Pennsylvania defendants and 70% of Bexar County, Texas defendants were never charged with felonies across their entire available criminal histories. Authors' calculations from the Administrative Office of Pennsylvania Courts data obtained by Crystal Yang via a public access request for Dobbie et al. (2018), and from Bexar County, Texas data used in Agan et al. (2021).

have less to lose from engaging in criminal activity, relative to nonprosecuted defendants who still have clean records to protect. District attorneys around the country are struggling with the policy question of misdemeanor prosecution: some have implemented presumptions of nonprosecution for certain nonviolent misdemeanor offenses, while others continue to advance all misdemeanor arrests to prosecution. The net causal effect of prosecution in marginal misdemeanor cases is an empirical question, but there is little evidence to guide prosecutors' policy choices.

In this paper we use new data on the prosecution of nonviolent misdemeanor criminal complaints from the Suffolk County District Attorney's Office (SCDAO) in Massachusetts between 2004 and 2018 to estimate the causal impact of nonprosecution on defendants' subsequent criminal justice involvement. Our empirical strategy exploits the as-if random assignment of cases to arraigning ADAs who vary in the leniency of their prosecution decisions. This empirical design recovers the local average treatment effect (LATE), or the causal effect of nonviolent misdemeanor nonprosecution for individuals at the margin of nonprosecution (Imbens and Angrist 1994; Doyle Jr 2007; Dahl et al. 2014; Dobbie et al. 2018; see Section 3 for a discussion of recent work by Blandhol et al. 2022 and Goldsmith-Pinkham et al. 2022).

We find that the marginal nonprosecuted misdemeanor defendant is 29 percentage points less likely to be issued a new criminal complaint (53% less than the mean for "complier" defendants who are prosecuted), and is issued 1.7 fewer complaints (60% fewer), within two years post-arraignment. Our results remain consistent through six years post-arraignment (if anything, impacts appear to grow over time). The results of our analysis imply that if all arraigning ADAs acted more like the most lenient ADAs in our sample when deciding which cases to prosecute, Suffolk County would likely see a reduction in criminal justice involvement for these nonviolent misdemeanor defendants. Approximately 46% of nonviolent misdemeanor defendants in Suffolk County are Black, while only approximately 24% of Suffolk County residents are Black (U.S. Census Bureau, 2019 Population Estimates Program).

Reducing the prosecution of nonviolent misdemeanor offenses would thus disproportionately benefit Black residents of the county.

Our treatment is whether an arraigning ADA does or does not advance a criminal complaint to prosecution. Defendants whose complaints are not advanced to prosecution have no further formal interactions with the district attorney's office and no additional case records beyond the day of arraignment. Defendants whose cases are advanced to prosecution may have a variety of additional interactions with the district attorney's office and additional case records beyond the day of arraignment. We use these additional records for prosecuted defendants to consider possible causal mechanisms that could be generating our findings, showing that neither the disruption caused by a lengthy prosecution nor the probability of an eventual misdemeanor conviction is likely driving causal effects.

Instead, it is the statewide criminal records of misdemeanor charges acquired by all prosecuted defendants that appear to be driving our findings. Consistent with this hypothesis, the proportional effects of nonprosecution are even larger for marginal defendants without prior criminal records (81% decrease in the probability of a new criminal complaint within two years post-arraignment), while the impacts for defendants who already have criminal records are noisy, statistically insignificant, and even suggest possible increases in new criminal complaints. Effects appear within the first three months post-arraignment, consistent with a behavioral change by prosecuted defendants in response to criminal record acquisition at arraignment. Nonprosecution has larger effects on subsequent arrests over which police officers have less discretion, consistent with changes in defendant criminal behavior, not only changes in law enforcement responses to the criminal records acquired by prosecuted defendants.

Our data are sourced from the internal case management records of the Suffolk County (Massachusetts) District Attorney's Office (SCDAO). These data have the advantage of recording information on criminal complaints that were not prosecuted and do not appear in court records. The SCDAO data also record information about each case event along with the identity of the arraigning ADA for at least some cases. One drawback of these administrative data is that the identity of the ADA at arraignment is missing for 67% of nonviolent misdemeanor cases meeting all other sample restrictions. Our main analysis is done within the sample of cases not missing this ADA information. However, we show that OLS estimates of the relationship between nonprosecution and subsequent criminal justice contact are nearly identical within the sample of cases missing arraigning ADA information. We further show that our main 2SLS results are quite similar within samples restricted to courts and years missing less ADA information, and within several progressively larger samples for which missing arraigning ADA information has been imputed based on patterns of observed arraigning ADAs by court/day and court/week. These results, together with our qualitative interviews of SCDAO staff, support our conclusion that arraigning ADA information is missing "as good as" randomly. Defendant race and ethnicity information is also sometimes missing (for 27% of nonprosecuted cases and 13% of prosecuted cases); we impute race/ethnicity using other defendant information.

Our 2SLS results are relevant for the approximately 10% of our sample that we identify as compliers. We explore two ways of moving beyond this LATE estimate. First, we estimate marginal treatment effects (MTEs) to explore heterogeneity in the LATE (Heckman and Vytlacil 2005; Heckman et al. 2006). We find no evidence that nonprosecution would increase recidivism for marginal defendants as leniency increases, though interpreting these estimates requires stricter assumptions than our LATE estimates, warranting caution. We also analyze a policy change in Suffolk County occasioned by the arrival of a newly elected district attorney who had campaigned on a platform of presumptive nonprosecution for nonviolent misdemeanor offenses. In a series of event study, difference-in-differences (DD), and 2SLS DD models using the date of the policy change as an instrument for nonprosecution, we find results that are consistent with our main results: increasing rates of nonprosecution for nonviolent misdemeanor cases reduced the likelihood of subsequent criminal complaints within a one-year post-arraignment window. In addition, there does not appear to have been an increase in reported crime due to the policy change.

Existing empirical work provides little guidance on the potential impacts of misdemeanor prosecution.[2] Most work on specific deterrence has focused on incarceration (which also imposes an incapacitation effect), and the findings are mixed. Some have found that incarceration or longer periods of incarceration decreases future crime; others have found increases in future crime (see Nagin et al. 2009; Raphael and Stoll 2014, Chalfin and McCrary 2017; and Doleac 2020 for reviews). Because misdemeanor prosecutions rarely result in incarceration sentences (in our sample, only 3.3% of prosecuted misdemeanor defendants receive incarceration sentences), these findings are of limited relevance. There is also evidence that sanctions (or more severe sanctions) for driving violations or DUIs decrease subsequent infractions for individuals who experience the sanction (Hansen 2015; Gehrsitz 2017; Dusek and Traxler 2020). However, driving violations do not generally result in criminal record acquisition, limiting the relevance of these findings to the context of misdemeanor prosecution.

There is also a small literature on the impact of diversion in the criminal justice system. In particular, Mueller-Smith and Schnepel (2021) study the impact of felony diversion in Texas, both deferred adjudication leading to dismissal of charges after completion of a period of probation and outright dismissal of charges, finding that marginal first-time felony defendants who received diversion (avoiding criminal records of felony conviction) had significantly lower probabilities of subsequent conviction and higher probabilities of subsequent employment. Criminal records of felony conviction may have different consequences for defendants, relative to criminal records of misdemeanor charges, limiting the applicability of Mueller-Smith and Schnepel (2021) to the question of misdemeanor prosecution. Augustine et al. (2021) study a post-charging felony diversion program in San Francisco, finding that felony diversion in a sample of defendants who all received criminal records of felony charges decreased the

[2] A few other studies consider variation in prosecutorial discretion. Rehavi and Starr (2014) and Tuttle (2021) both report evidence that federal prosecutors exhibit racial bias in their prosecution decisions. Sloan (2020a) and Sloan (2020b) use random assignment of cases to prosecutors in the office of the District Attorney of New York County to document variation in prosecutorial leniency and to test for other-race bias in prosecutors' decisions. However, none of these studies estimates the causal impacts of prosecutors' decisions on misdemeanor defendants' subsequent outcomes.

probability of subsequent conviction. Finally, Rempel et al. (2018) use a matching design to compare outcomes for defendants who did and did not receive pretrial diversion in five jurisdictions, including pre- and post-charge diversion programs for both misdemeanor and felony defendants, finding in four jurisdictions that diversion reduced rearrest.

It is unclear ex ante if the downstream effects of misdemeanor prosecution are more likely to be similar to those that follow from driving infractions, or to those that follow from felony conviction. The net effect of misdemeanor prosecution is an empirical question—one that is being hotly debated in many cities and counties around the United States.

This study contributes to the literature in several ways. First, we provide evidence on the causal effects of the decision to prosecute a nonviolent misdemeanor defendant. Given that misdemeanors make up the vast majority of charges in the criminal justice system, and that many defendants will never experience a felony charge, this is an important policy lever. District attorneys across the country are implementing policies with a presumption of non-prosecution for subsets of nonviolent misdemeanor offenses, making this decision importantly policy relevant.

Second, because prosecution in our setting determines whether a defendant acquires a criminal record of misdemeanor charges, our findings also speak to the consequences of the "mark of a criminal record" (Pager 2003). While much of the research on this question has examined the consequences of felony conviction records, there is some evidence on the consequences of misdemeanor records. Field experimental research has found that employers are less likely to call back individuals with misdemeanor conviction records (Leasure 2019) or even only nonviolent misdemeanor arrest records (Uggen et al. 2014). Smith and Broege (2020) likewise find that individuals with criminal legal contact of any kind (not just convictions) are less likely to search for jobs post-contact than individuals who were otherwise similar pre-contact. Our findings contribute to this literature on the consequences of misdemeanor criminal records.

Third, our findings contribute to the literature on the net costs and benefits of criminal justice intervention, and on the diminishing marginal returns to such interventions. In this context, it appears that prosecuting defendants for nonviolent misdemeanor offenses has substantial costs for those individuals without any evidence of public safety benefits (and suggestive evidence of public safety costs).

Finally, we add to a growing literature that uses as-if randomization of cases to decision makers (in this case arraigning ADAs) to measure the causal effects of their decisions. There has been substantial recent work refining this econometric method, and we apply this method in a new context, using current best practices.

2 Setting and Data

In this paper we study the effects of nonviolent misdemeanor prosecution in Suffolk County, Massachusetts (which includes the cities of Boston, Chelsea, Revere, and Winthrop). Our data are sourced from the internal case management records of the Suffolk County District Attorney's Office, and include a record of all criminal complaints issued in the county between January 1, 2000 and September 1, 2020, including complaints that were not prosecuted and thus do not appear in court records.[3]

In Suffolk County, misdemeanor complaints are processed in one of nine municipal or district courts, each of which has a geographically defined jurisdiction. Misdemeanor criminal complaints are calendared for arraignment by the court with geographic jurisdiction over the location at which the alleged offense occurred. Assistant District Attorneys (ADAs) are scheduled to arraignment courtrooms in each of the nine municipal and district courts by SCDAO supervisors on an approximately weekly basis based on availability. ADAs assigned to an arraignment courtroom on a given day are responsible for arraigning all of the cases

[3] In Massachusetts, clerk magistrates in the courts of jurisdiction review police officers' criminal complaint "applications" for completeness, calendar complaints for arraignment, issue summonses for defendants to appear at arraignment (if defendants are not already in custody), and provide arraigning ADAs with complaint records at the beginning of arraignment sessions. Our interviews with SCDAO staff indicate that as long as there is a police report attached to an application averring that the defendant committed a criminal offense, the complaint is calendared; our main impacts are estimated for the sample of complaints calendared for arraignment and provided to arraigning ADAs.

calendared for that courtroom on that day. Absent a conflict of interest in a specific case (e.g., the arraigning ADA went to school with the defendant), defendants may not request a different arraigning ADA, nor may arraigning ADAs choose which cases to arraign. There are a few exceptions to this general practice, triggered by specific charge types. Cases with felony charges may receive additional scrutiny from supervising ADAs, strategic assignment to more experienced arraigning ADAs, and/or the involvement of ADAs from the Superior Court. We therefore exclude any cases in which defendants are charged with felony offenses, regardless of the final disposition of those felony charges. We also exclude cases with violent or firearm charges for similar reasons: in these cases more experienced ADAs may be called in to support or handle the arraignment.

In our data, ADAs are assigned to arraignment courtrooms for an average of 85 days dispersed across an average of 3.4 years.[4] For cases that proceed past the day of arraignment, a second and separate SCDAO procedure assigns an ADA to oversee all subsequent case stages. All other court actors, such as judges and public defenders, are also assigned to cases through procedures that are independent of the process through which arraigning ADAs are assigned to arraignment courtrooms.

2.1 Defining Treatment

During an arraignment hearing a defendant's name and charges are read into the record by the court clerk and the defendant enters a plea before the arraigning judge. Under Massachusetts law, a complaint that does not proceed to this formal arraignment hearing (including, critically, the reading of the defendant's name and charges into the court record) is not recorded in the statewide criminal records maintained by the Massachusetts Department of Criminal Justice Information Services (DCJIS).[5] Complaints that do proceed to the

[4] This undercounts the duration of ADA arraignment rotations because the count is based only on the cases for which arraigning ADA information is recorded. Section 4.3 explores a variety of strategies to address missing ADA information.

[5] https://malegislature.gov/Laws/GeneralLaws/PartI/TitleII/Chapter6/section167: "Such information ["criminal offender record information"] shall be restricted to information recorded in criminal

moment of formal arraignment are recorded in the DCJIS database, and are available to employers under certain conditions upon request, even if the defendant is not convicted on any charges.[6]

In practice, an arraigning ADA is given a large stack of paper files in the arraignment courtroom on the morning of an arraignment shift, and needs to quickly work through how to proceed in each case. The arraigning ADA has the discretion to not advance a complaint to the formal moment of arraignment, or to proceed with arraignment. If the arraigning ADA chooses to advance a complaint to the moment of formal arraignment, the defendant's name and charges will be read into the court record, and the defendant will acquire a criminal record of charges in the statewide DCJIS database.

During the formal arraignment hearing, a defendant may plead guilty, not guilty, offer an "admission to sufficient facts" leading to pretrial probation, or request a two-week arraignment continuance to be considered for diversion leading to dismissal of charges.[7] Arraigning ADAs may also choose to dismiss defendants' charges on the day of arraignment, after the formal moment of arraignment.

In our data, we observe all case events, the dates of those events, and final charge dispositions. We define "nonprosecution" as cases with no further case events after the day of the initial arraignment, and no dispositions recorded as a conviction or an admission to sufficient facts; "prosecution" includes all other cases. Given the nature of our data, we cannot further distinguish between cases that are not advanced to the moment of arraignment, and cases that are dismissed on the day of arraignment after the formal arraignment hearing. However, we can show that cases that we define as "nonprosecution" are much less likely to be recorded in the Massachusetts DCJIS criminal records data, relative to cases that we define as "prosecution."

---

proceedings that are not dismissed before arraignment."

[6] https://www.mass.gov/service-details/levels-of-name-based-criminal-record-check-access.

[7] https://malegislature.gov/Laws/GeneralLaws/PartIV/TitleII/Chapter278/Section18; https://malegislature.gov/laws/generallaws/partiv/titleii/chapter276a.

2.2 Sample

A complaint can contain multiple arrest charges; we refer to each complaint as a case. Cases are dated using the date of the first "event" recorded in a case; we refer to that date as the day of arraignment. Case records include an identifier for the court of jurisdiction; we exclude cases brought in a court other than one of the nine municipal/district courts. Defendants are identified with unique IDs, enabling us to link cases across defendants within Suffolk County. In our main analysis, we follow each defendant in a case for a period of two years following arraignment, including cases with arraignment dates between January 1, 2004 and September 1, 2018. We use data from January 1, 2000 to generate criminal histories, and we follow defendants up to September 1, 2020. In supplemental results we analyze one- to six-year followup periods; the one-year followup sample adds one additional year of criminal cases; the longer followup samples subtract one year of criminal cases for each additional followup year.

98.5% of charges are identified in the SCDAO data with an offense severity code indicating whether a charge is a misdemeanor, a felony, or a civil violation (e.g., a civil motor vehicle violation). We exclude any case with at least one charge identified as a felony charge. We use text extraction to identify charge types. As described previously, violent offenses may be treated differently during arraignment and thus we exclude cases with any charge for a violent offense—including assault, assault and battery, violating a domestic abuse prevention order, and criminal harassment—and those with any firearms-related charges. We sort the remaining charges into the following categories: motor vehicle, drug, disorder/theft, and other. We refer to this final set of charges as "nonviolent misdemeanors."

Charges are associated with a variety of different final disposition codes. We characterize final dispositions at the charge level as resulting in a criminal conviction or no conviction. Final dispositions that result in convictions are pleas of guilty and guilty verdicts after bench or jury trials. Final dispositions that do not result in conviction are all other dispositions, including dismissal, pretrial probation, nolle prosequi, admission to sufficient facts, or a finding of not guilty after a jury or bench trial.

Arraigning ADAs are identified in the SCDAO data for 33% of nonviolent misdemeanor cases in our sample arraigned between 2004 and 2018. Our main analysis is done within the sample of cases not missing arraigning ADA information. In Section 4.3 and Appendix C.2. we explore the missingness of ADA information, finding that the missingness of arraigning ADA is unrelated to other case and defendant features, and that our findings are robust to estimation within subsamples with less missing data and several strategies for imputing missing data on arraigning ADAs.

Defendant sex and age are missing for 1.4% of observations; we exclude these observations from our sample. We sort defendants into age groups representing the 25th, 50th, and 75th percentiles of age to allow for measurement error in precise age. As we report in Appendix C.1, defendant race/ethnicity was systematically less likely to have been recorded both for defendants who were not prosecuted, and for defendants who were not rearrested during our time period. We therefore predict race/ethnicity using the procedures reported in Appendix C.1, and include indicators for whether a defendant is most likely to be Hispanic, Black, or white as covariates in all analyses. In Appendix Table C.2 we show that within the sample for which we do have administrative race data, imputed race is highly correlated with administrative data on recorded race.

Our main estimation sample includes cases whose arraignment hearings occur between January 1, 2004 and September 1, 2018; that do not include violent, firearms, or felony charges; that are arraigned in one of Suffolk County's nine district/municipal courts; and for which arraigning ADA, sex, and age information are not missing. We further restrict our estimation sample to those nonviolent misdemeanor cases overseen at arraignment by an ADA who oversees at least 30 other nonviolent misdemeanor cases at arraignment hearings, and to those cases that are not "singletons" within our set of court-by-time fixed effects (defined below).

SCDAO case records were matched by docket number to the criminal records database maintained by the Massachusetts DCJIS. Not all SCDAO case records matched to the DCJIS database. Cases that are disposed of prior to arraignment do not result in DCJIS records. Other SCDAO case records may not match to a DCJIS case record because of human error in docket number entry.

2.3 Descriptive Statistics

Table 1 reports descriptive statistics for this sample. There are 67,060 cases in the SCDAO data that meet these criteria. Using our definition of prosecution, 20% of these nonviolent misdemeanor cases are not prosecuted; the remaining 80% are prosecuted. 73% of nonviolent misdemeanor cases that are prosecuted are eventually disposed of without criminal convictions.

Nonviolent misdemeanor cases that are not prosecuted are clearly different from cases that are prosecuted. Nonviolent misdemeanor defendants who are not prosecuted are issued criminal complaints that include fewer counts overall and fewer "serious" misdemeanor counts (punishable by greater than 100 days incarceration). They are less likely to have had a misdemeanor or felony conviction within one year prior to the arraignment hearing in their case. They are more likely to be citizens, female, and (predicted) white. They are more likely to have been charged with a motor vehicle offense, and less likely to have been charged with a drug offense or a disorder/theft offense. For the purpose of assessing the monotonicity of our instrument, we also code offense types as "victimless" or "victim" offenses. "Victim" offenses include property offenses (e.g., larceny, shoplifting, burglary), threats, property damage, and leaving the scene of property damage or personal injury. Defendants who are not prosecuted are more likely to have been charged with a "victimless" offense.

By construction, defendants who are not prosecuted have fewer days to disposition, fewer case events, and are less likely to receive convictions, relative to prosecuted defendants. They are also less likely to acquire DCJIS records of their complaint, relative to prosecuted defendants. Defendants who are not prosecuted are then less likely to receive a new criminal complaint within two years, relative to defendants who are prosecuted. However, because prosecution is clearly correlated with observable pre-treatment characteristics (and likely correlated with unobservable pre-treatment characteristics as well), we cannot draw conclusions about the effect of prosecution on the probability of post-arraignment outcomes from these data alone.

3 Research Design

We want to estimate the effect of misdemeanor nonprosecution on post-arraignment outcomes. Consider the following model, where Y_{ict} is the outcome of interest for individual i in case c in year t, $\mathbf{X}_{ict}$ is a vector of case- and defendant-level covariates, γ_{ct} are court-by-time fixed effects described later, and ε_{ict} is an error term:

$$Y_{ict} = \beta_1 \mathit{Not\ Prosecuted}_{ict} + \beta_2 \mathbf{X}_{it} + \gamma_{ct} + \varepsilon_{ict} \tag{1}$$

β_1 is our parameter of interest. The key problem for causal inference is that ordinary least squares (OLS) estimates of Equation 1 are likely to be biased by the correlation between prosecution and unobserved defendant characteristics that are correlated with outcomes. This selection bias could be either positive or negative. For example, arraigning ADAs are more likely to prosecute misdemeanor defendants who have prior criminal convictions, and defendants with prior convictions are also more likely to have subsequent criminal justice contact. Arraigning ADAs are less likely to prosecute younger defendants, and younger defendants are also more likely to have subsequent criminal justice contact (see Appendix Table B.1). Unobservable characteristics could cause selection bias in either direction as well.

The as-if random assignment of misdemeanor cases to arraigning ADAs creates the opportunity to identify a source of variation in nonprosecution that does not depend on defendant or case characteristics. We estimate the causal impacts of misdemeanor nonprosecution by using the propensity of an as-if randomly assigned ADA to not prosecute a defendant as an instrument for nonprosecution.

We construct a residualized leave-out ADA leniency measure for our instrument (French and Song 2014; Dahl et al. 2014; Dobbie et al. 2018).Because misdemeanor case types may vary by court, year-month, and day of week, a simple leave-out measure of ADA leniency could be confounded by selection. To address this, we include court-by-year-month and court-by-day-of-week fixed effects, γ_{ct}, in the construction of our instrument. The inclusion of these court-by-time fixed effects allows us to interpret variation in the instrument as variation in the tendency of an as-if-randomly assigned ADA to prosecute a nonviolent misdemeanor defendant, relative to the other nonviolent misdemeanor cases brought in that court in that year-month, and in that court on that day of the week.[8] Call this residual nonprosecution decision $Not\ Prosecuted^{*}_{ict}$.

As is standard to avoid the small-sample correlation between the ADA decision in this case and her average leniency, we then construct the leave-out mean measure of ADA nonprosecution (leniency) for each nonviolent misdemeanor case using these residual nonprosecution decisions:

$$\mathrm{Z}_{cta} = \left(\frac{1}{n_a - n_{ia}}\right)\left(\sum_{k=0}^{n_a}(Not\ Prosecuted^{*}_{ikt}) - \sum_{c=0}^{n_{ia}}(Not\ Prosecuted^{*}_{ict})\right) \tag{2}$$

where n_a is the number of nonviolent misdemeanor cases arraigned by ADA a and n_{ia} is the number of nonviolent misdemeanor cases involving defendant i arraigned by ADA a. This construction removes from the instrument the residualized nonprosecution decisions in all of a defendant's nonviolent misdemeanor cases arraigned by ADA a.

Figure 1 reports the distribution of our residualized ADA nonprosecution measure. As noted previously, we restrict the sample to exclude nonviolent misdemeanor cases overseen

[8] In Appendix Table B.9 we also consider a version of the instrument that uses court x week rather than court x month fixed effects, and a "raw" measure of ADA leniency based on the non-residualized nonprosecution rate, and find similar results in both cases.

by arraigning ADAs assigned to fewer than 30 nonviolent misdemeanor cases, and cases that are "singletons" within our set of court-by-time fixed effects. After these restrictions, the sample includes 315 arraigning ADAs. The median number of nonviolent misdemeanor cases overseen by an arraigning ADA is 155 cases; the average is 212 cases. After residualizing out our set of court-by-time effects, the ADA measure ranges from -0.08 at the first percentile to 0.09 at the ninety-ninth percentile. That is, moving from the first to the ninety-ninth percentile of ADA leniency increases the rate of nonprosecution by 17 percentage points, an 83% change from the mean nonprosecution rate of 20.4%. In Appendix Figure B.1 we also show this variation after applying an empirical Bayes shrinkage procedure to adjust for sampling error and still see dispersion in our leniency measure (see Appendix A.3 for details).

Our main analysis will be based on 2SLS estimates of second-stage Equation 1 (with and without case- and defendant-level covariates) and a first stage for individual i and case c assigned to arraigning ADA a at time t, using a linear probability model:

$$\textit{Not Prosecuted}_{icta} = \alpha_1 Z_{cta} + \alpha_2 \mathbf{X}_{it} + \gamma_{ct} + \varepsilon_{ict} \tag{3}$$

where, again, γ_{ct} are the court-by-year-month and court-by-day-of-week fixed effects, and $\mathbf{X}_{ict}$ includes case- and defendant-level covariates (number of counts, number of misdemeanor counts, number of serious misdemeanor counts, any convictions for felonies or misdemeanors in the previous year, offense type, citizenship, gender, age, and predicted race/ethnicity). Z_{cta} are the leave-out measures of residualized ADA leniency described previously (in Section 4.2 we also consider alternative ways of constructing the instrument). Robust standard errors are clustered at both the defendant and ADA level. We report the robust first-stage F-statistic, which is large in our setting (Montiel Olea and Pflueger, 2013). Rather than rely on a threshold rule based on this first-stage F-statistic, we also construct Anderson-Rubin confidence intervals, which are of correct size and optimal power even with weak instruments when treating the leniency measure as a single non-constructed instrument (Anderson and Rubin 1949; Andrews et al. 2019; Lee et al. 2020; for more on these confidence intervals in over-identified models see Davidson and MacKinnon 2014).[9]

We interpret our 2SLS effects in the local average treatment effect (LATE) framework (Imbens and Angrist 1994). That is, if the assumptions discussed below hold, we are able to recover the local causal effects of misdemeanor nonprosecution decisions for defendants on the margin of being not prosecuted—those whose treatment status would be changed by switching from a less to a more "lenient" ADA at arraignment. We note that the as-if randomization and relevance assumptions do not rely on the case- and defendant-level covariates X but do rely on the court-by-time fixed effects, which enter the first and second stages nonparametrically, thus avoiding the main concern in Blandhol et al. (2022). Appendix A.1 also includes a "saturated and weighted" specification which uses the individual ADA interacted with the court-by-time fixed effects as instruments (Angrist and Imbens 1995, Angrist and Pischke 2009 Blandhol et al. 2022), a specification also suggested by Goldsmith-Pinkham et al. 2022 to address potential contamination bias from many treatments and flexible controls in the first stage; the results remain similar. We argue that our LATE estimate is also a policy relevant treatment effect (PRTE)—it estimates the local effect of policies that increase the leniency of prosecution decisions for marginal defendants (Heckman and Vytlacil 2001).

In addition, we also estimate marginal treatment effects (MTEs) to explore heterogeneity based on unobservables and to understand the distribution of treatment effects (Björklund and Moffitt 1987; Heckman and Vytlacil 2005; Heckman et al. 2006). The MTEs are the average effects of nonprosecution for defendants on the margin between being prosecuted and not prosecuted, where these margins correspond to percentiles of the distribution of the unobserved resistance to being not prosecuted. Estimating the MTEs requires the same assumptions as the LATE framework, including strict monotonicity, plus the additional assumption that there is additive separability between the observed and unobserved heterogeneity in the

[9] It is somewhat of an open question how to evaluate the possibility of many-weak-instrument bias in leniency/examiner designs (Hull 2017; Frandsen et al. 2019; Bhuller et al. 2020). In Appendix A.1 we further explore alternative IV specifications that account for potential biases from the construction of our leniency measure, including: using all the ADA dummies directly as instruments (in a standard 2SLS setup and using limited information maximum likelihood estimation (LIML)), using lasso to pick the most informative ADA dummies, and using the UJIVE estimation strategy proposed by Kolesár (2013).

treatment effects, needed when the propensity score does not have full support, as ours does not (see e.g. Brinch et al. 2017, Mogstad and Torgovitsky 2018, Andresen 2018). These are strong assumptions, and thus we see the MTE estimates as suggestive. For further details on the derivation of the MTEs in the potential outcomes framework see Appendix A.4.

3.1 Assessing the Instrument

3.1.1 Exogeneity

In order to be able to interpret our 2SLS estimates as the local average treatment effect (LATE) of misdemeanor nonprosecution, it must be the case that defendant and case characteristics do not covary systematically with arraigning ADA assignment. Appendix Table B.1 reports the results of this randomization test. Column (1) reports linear probability estimates of the correlation between nonprosecution and case and defendant characteristics, after controlling for court-by-time fixed effects and clustering standard errors at both the defendant and the ADA level. Mirroring what we saw in the summary statistics, even with court-by-time fixed effects we see that the decision to not prosecute a particular defendant is highly correlated with defendant/case characteristics. Column (2) uses our residualized ADA leniency instrument as the dependent variable instead. With only one exception out of 17 coefficients (defendant gender), our measure of ADA leniency is not correlated with these observable characteristics. Consistent with our understanding that cases are allocated as-if randomly to arraigning ADAs, arraigning ADAs with varying propensities to prosecute handle very similar nonviolent misdemeanor cases (joint p-value = 0.17).

3.1.2 Instrument Relevance (First Stage)

Table 2 reports first stage results from Equation 3. Consistent with Figure 1, our residualized ADA instrument is highly predictive of whether a defendant is not prosecuted. The estimated first stage result is robust to the inclusion of controls in column (2), which is consistent with as-if random arraigning ADA assignment. With all controls, a nonviolent misdemeanor defendant assigned to an arraigning ADA who is 10 percentage points more likely to not prosecute a defendant is approximately 5.4 percentage points more likely to be not prosecuted.[10]

3.1.3 Exclusion Restriction

The exclusion restriction requires that arraigning ADAs only systematically affect defendant outcomes through the prosecution decision. We cannot directly test this assumption. However, we believe that the exclusion restriction assumption is reasonable in our setting. First, ADAs are almost exclusively only making a decision about prosecution at the arraignment hearing. Bail decisions for prosecuted defendants may also occur at arraignment but are rare for the nonviolent misdemeanor cases we study (see Section 4.2), implying that multidimensional decision-making is not an issue for arraigning ADA decisions. Second, for nonprosecuted defendants there are no further interactions with the district attorney's office and thus no scope for any correlations between arraignment ADA leniency and later outcomes. For prosecuted defendants, recall that after arraignment a different process is used to assign an ADA to take over subsequent case stages. All other court actors, such as judges and public defenders, are also attached to cases through a different process. These institutional characteristics make it unlikely that the assignment of an arraigning ADA is correlated with post-arraignment actions that could independently affect prosecuted defendant outcomes.

Appendix Table B.4, Panel A reports estimates of the association between arraigning ADA leniency and later case outcomes for prosecuted defendants only. Within the sample of cases that are prosecuted, there are no associations between arraigning ADA leniency and the number of case events, the number of days from arraignment to disposition, and the probability that a defendant receives a conviction. More lenient arraigning ADAs are

[10] This table also reports robust first-stage 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 the critical value of 23.11 they propose for just-identified models with τ =10% of worst case bias.

weakly associated with a slightly lower probability that bail is set for a prosecuted defendant at arraignment. However, because our sample comprises relatively minor offenses, bail is typically not requested: the arraigning ADA requests bail in only 8% of cases that they choose to prosecute, and bail is set by the judge in only 6.6% of prosecuted cases.[11] In Section 4.2 we explore the impact of arraigning ADAs' bail requests on our results, showing that the only meaningful channel through which arraigning ADAs affect defendants' outcomes is through the prosecution decision.

It is also not clear that if we did find associations between ADA leniency and case outcomes, this would represent a violation of the exclusion restriction. For example, if arraigning ADAs made prosecution decisions based only on the probability of conviction in a given case, and more lenient ADAs had a higher probability threshold for prosecution, then we would expect to see higher conviction rates for those defendants assigned to more lenient arraigning ADAs. This correlation would be due to selection, and would not constitute a violation of exclusion. That we do not see this correlation may tell us something about prosecutor objective functions, to which we return in Section 4.1.

3.1.4 Monotonicity

Under heterogeneous treatment effects, in order to recover the LATE—the causal impact for the compliers—we also need it to be the case that the impact of ADA assignment on the probability of nonprosecution is monotonic across defendants. This monotonicity assumption implies that defendants who are not prosecuted by stricter ADAs would also be not prosecuted by more lenient ADAs, and that defendants prosecuted by more lenient ADAs would also be prosecuted by stricter ADAs.

We cannot test this assumption directly, but there are several indirect tests we can pursue. Frandsen et al. (2019) show that one can relax the strict (pair-wise) monotonicity

[11] Using data from Massachusetts Trial Courts and DCJIS, Bishop et al. (2020) show that across all cases—including violent misdemeanors and felonies—bail is imposed in 11-17% of arraignment hearings depending on the race of the defendant. Given that our analysis focuses on nonviolent misdemeanors, the low rates of bail assigned at arraignment we observe appear consistent with statewide data.

assumption of the original LATE framework to an average monotonicity assumption and still recover a weighted average of individual treatment effects. This average monotonicity assumption implies that the covariance between a defendant's prosecutor-specific treatment and the prosecutor's overall propensity to not prosecute is weakly positive. One test of this assumption is that prosecutors' group-specific nonprosecution rates should be positively correlated with overall nonprosecution—prosecutors who are more lenient overall should be more likely to not prosecute people in any observable subgroup. This is equivalent to showing that the first stage should be positive in all subsamples of the data, as is common in the literature (Dobbie et al. 2018; Bhuller et al. 2020). Appendix Table B.2 presents first stage results for a large variety of subsamples of our data. Consistent with the average monotonicity assumption, we find that the relationship between our residualized measure of ADA leniency and nonprosecution is positive and significant in all subsamples. In specification checks in Section 4.2, we also create versions of our instrument that are interacted with various ADA and case characteristics to relax these monotonicity assumptions. Frandsen et al. (2019) also provide a test for the joint null hypothesis that the exclusion and monotonicity assumptions hold. We calculate this test within the nine courts in our dataset, controlling for our main set of covariates and year-month and day-of-week fixed effects. In Appendix Table B.3 we show that within six of the nine courts in our data we fail to reject the joint null hypothesis that exclusion and monotonicity hold.[12]

[12] de Chaisemartin (2017) offers another way of relaxing the strict monotonicity assumption. Under a weaker condition he calls the "compliers-defiers" (CD) condition—for any pair of ADAs, there is a subset of compliers that is the same size as the subset of defiers (defendants who violate monotonicity for this pair) and that has the same local average treatment effect as the defiers—the 2SLS estimates are the LATE for the subgroup of the remaining compliers. The CD condition holds if the treatment effect has the same sign for both compliers and defiers, and if the treatment effect for compliers is greater than the treatment effect for defiers. We do not have strong reasons to believe that compliers and defiers would have differently signed treatment effects. This weaker "compliers-defiers" condition is also tested by the joint monotonicity-exclusion test of Frandsen et al. (2019), which we fail to reject across a large share of our sample.

4 Results

Table 3 reports OLS and 2SLS estimates of the impacts of nonviolent misdemeanor non-prosecution on the likelihood and number of subsequent criminal complaints within two years post-arraignment, as well as reduced form estimates of the effect of ADA leniency on the likelihood of subsequent complaints. OLS estimates with controls in Panel A, column (2) imply that nonprosecution reduces the probability of a subsequent criminal complaint by 10 percentage points (a 27% decrease relative to the mean for prosecuted defendants). The 2SLS estimates with controls in Panel A, column (4) indicate that nonprosecution of marginal defendants reduces the probability of a subsequent criminal complaint by 29 percentage points ($p < 0.01$), a 53% decrease relative to the mean for prosecuted complier defendants.[13] Reduced form estimates in Panel A, column (5) are also large, negative, and statistically significant. If the exclusion restriction is violated, reduced form estimates can still be interpreted as the causal effects of being assigned to a more or less lenient arraigning ADA. Appendix Figure B.2 shows a non-parametric version of the reduced form, mimicking Figure 1, showing that estimates are negative across the distribution of ADA leniency. In Panel B we consider the intensive margin of future complaints, finding that nonprosecution reduces the number of subsequent criminal complaints for marginal defendants by 1.7 complaints (60%; $p < 0.05$).

Figure 2 shows how the effect of nonprosecution evolves over the two-year followup period in three month bins. We see an immediate drop in the likelihood of a new complaint within the first three months after arraignment, and this effect remains steady through the first post-arraignment year. At that point the negative effect begins to grow larger over time. The effect of nonprosecution is not just a short-run phenomenon. Figure 3 reports 2SLS estimates with covariates for various time horizons from one to six years post-arraignment. Even six years post-arraignment, marginal non-prosecuted defendants are significantly less

[13] See Appendix A.2 for details on the calculation of average outcomes among prosecuted compliers.

likely to receive new criminal complaints than prosecuted defendants (36 pp, a 55% decline over the control complier mean). Appendix Table B.5 replicates Panel A of Table 3 for one-, three-, and five-year time horizons post-arraignment. For the remainder of the paper we will focus robustness and subsample results on the two-year post-arraignment window, although results remain similar, if not larger, for longer post-arraignment horizons.

Appendix Table B.6 reports 2SLS estimates with all controls for complaints within two years by subsequent crime type (violent, motor vehicle, disorder/theft, drug, and other) and subsequent offense seriousness (misdemeanor/felony). We find significant decreases in subsequent violent and disorder/theft charges, both overall and for subsequent misdemeanor charges. Nonprosecution reduces the rates at which nonviolent misdemeanor defendants are charged with subsequent violent offenses by 65%, and with subsequent disorder/property offenses by 83%. Appendix Table B.7 reports heterogeneity across demographic groups. Across subgroups 2SLS estimates are negative, although not always statistically significant at conventional levels. There do not appear to be meaningful differences across gender or predicted race/ethnicity.

4.1 Selection and Compliers

Our 2SLS estimates represent the LATE for marginal defendants—defendants who would have received a different prosecution decision had their case been assigned to a different arraigning ADA. Our main OLS estimates are smaller in absolute value than our 2SLS estimates. OLS and 2SLS estimates can differ due to heterogeneity in the effect of non-prosecution on subsequent criminal justice contact for the compliers and/or due to selection bias.

We characterize the share of compliers and their characteristics following the approach developed by Abadie (2003) and Dahl et al. (2014), and applied by Dobbie et al. (2018) and Bhuller et al. (2020). Details of these calculations can be found in Appendix A.2. We estimate that around 10% of our sample are compliers. While compliers look similar to the full sample on some dimensions, they differ on others (see Appendix Tables B.10 and B.11).[14] To explore possible heterogeneity, in Appendix Table B.12 we reweight our OLS estimates to match the sample of compliers using two different reweighting schemes (Dahl et al. 2014; Bhuller et al. 2020). We see that the reweighted OLS estimates are very similar to the unweighted OLS estimates under both reweighting schemes, implying that the differences between the OLS and 2SLS estimates are unlikely to be accounted for by heterogeneity in causal effects for compliers by observable characteristics (we cannot rule out heterogeneity on unobservable characteristics).

The differences we see, then, are likely driven by selection bias: arraigning ADAs are, on average, choosing to not prosecute defendants who have higher risk of subsequent criminal justice contact than marginal defendants. This may at first glance seem counterintuitive. However, there are a variety of characteristics that ADAs might interpret as mitigating circumstances making defendants less culpable of their crimes or more worthy of a second chance, but that also increase the risk of subsequent criminal justice contact, for example mental health issues, drug addiction, or age. We can see in Appendix Table B.1, column (1) that older defendants are less likely to be not prosecuted than defendants age 23 or younger (the base group in that regression); that is, younger defendants are less likely to be prosecuted. However, younger defendants are at significantly higher risk of future criminal justice contact than are older defendants.[15] It is likely that other unobserved characteristics, including mental health issues and substance abuse, may induce a similar type of negative selection, moving arraigning ADAs to choose leniency for defendants that have higher risk of future criminal justice contact. We return to the question of the prosecutor's objective function in Section 6.1.

[14] In particular, compliers are less likely to have been charged with a drug offense, to have been charged with a serious misdemeanor (punishable by more than 100 days in jail), to have misdemeanor or felony convictions within the prior year, and to be noncitizens, and more likely to be younger (less than 24 years old) and female.

[15] It is true generally that younger people are more likely to have criminal justice contact, relative to older people (Laub and Sampson 2001; Landersø et al. 2017), and also true in our data.

4.2 Specification Checks

In this section we pursue a variety of modifications to our primary specifications to probe the robustness of our results. First, as discussed earlier, ADAs can make bail requests at the arraignment hearing (although for the nonviolent misdemeanor cases in our sample bail is requested in only 8% of cases). We might worry that our leniency measure confounds two types of leniency: "nonprosecution leniency" and "no-bail leniency." In Appendix Table B.8 we address this in three ways, showing the 2SLS estimate for the effect on subsequent criminal complaints in each case. First, we create a "no-bail leniency" measure based on ADAs' propensity to request bail in other defendants' cases, and simply control for it in our regressions. Our results are nearly identical. Second, we use our no-bail leniency measure as an instrument for not receiving bail, and estimate the effect on subsequent complaints. We find a negative coefficient (which could be due to the correlation of the bail decision with the nonprosecution decision) but it is insignificant. Third, we use both nonprosecution leniency and no-bail leniency as instruments in the same regression, to measure the separate effects of the nonprosecution and no-bail decisions. Our estimate for nonprosecution is nearly identical to our main estimate, and the estimated effect of no-bail is near-zero and statistically insignificant. Based on these results, we conclude that arraigning ADAs' decisions about whether to request bail do not explain our results. These analyses support our hypothesis that the only meaningful channel through which arraigning ADAs affect defendants' outcomes is through the decision of whether to prosecute the case.

Appendix Table B.9 reports 2SLS estimates of the probability of receiving a subsequent complaint within two years for different versions of our instrument. Panel B column (1) reports the 2SLS estimate using a version of our leave-out mean instrument that does not residualize out court-by-time fixed effects. This instrument is thus a raw measure of an ADA's leave-out nonprosecution rate. In Panel B column (2), we use empirical Bayes shrinkage to shrink our leniency estimates towards a prior mean of 0 (see Appendix A.3 for details).

Columns (3)-(5) report estimates for more flexible instruments constructed by interacting our main leave-out instrument with various ADA or case characteristics. This relaxes our monotonicity assumption and allows the effect of ADA leniency to vary with each of the following: (i) high versus low ADA experience (as measured by above- or below-median number of nonviolent misdemeanors arraigned as of the time of this case's arraignment), (ii) whether the crime is categorized as victimless, or (iii) several mutually-exclusive crime types. In all cases estimates are qualitatively similar to the main estimates presented above; coefficients maintain the same sign and are of similar magnitudes and significance.

In Appendix A.1 we further explore alternative IV specifications that account for potential biases from the construction of our leniency measure, including using all the ADA dummies directly as instruments, using limited information maximum likelihood estimation (LIML), using lasso to pick the most informative ADA dummies, and using the UJIVE estimation strategy proposed by Kolesár (2013). In Appendix Table A.1 we see that across different estimation strategies we robustly find a negative relationship similar in magnitude to our baseline estimate.

4.3 Missing Data

Our data have many advantages, including allowing us to see criminal complaints that, because they were not prosecuted, do not appear in court records. However, our data also have extensive missingness on the identity of the arraigning ADA, with 67% of cases meeting all other sample criteria missing information on arraigning ADA identity. Arraigning ADA information is entered into the office's electronic case management system from paper files by administrative assistants who are pressed for time, and who may not prioritize the electronic capture of arraigning ADA identity, particularly for cases that are not proceeding past arraignment. Arraigning ADA information is missing in 64% of cases that are prosecuted, and in 75% of cases that are not prosecuted. Further details on missing ADA data are available in Appendix C.2.

ADA missingness would bias our estimates if defendants in cases missing electronic arraigning ADA information were less likely to be be issued new criminal complaints after being prosecuted, and/or were more likely to be issued new criminal complaints after being not prosecuted. In that case, we would expect to see smaller negative (or even positive) OLS estimates of the effect of nonprosecution on recidivism within the sample of cases missing arraigning ADA information. Panel A of Table 6 reports OLS estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within two years for our main estimation sample (not missing ADA information), the sample of cases meeting other sample criteria but missing ADA information, and the combined sample ("full relevant sample"). The OLS estimates are very similar across cases that are missing and not missing arraigning ADA information. In addition, in Appendix Table C.4 we regress an indicator for missing ADA information on various case and defendant characteristics and subsequent criminal complaints within two years and see there is no correlation between ADA missingness and subsequent complaints for either prosecuted or not prosecuted defendants.

We then re-estimate our 2SLS models in samples with less missingness. Panel B, column (1) of Table 6 repeats the main 2SLS estimates from column (4) of Table 3. Column (2) of Panel B replicates this analysis within courts where arraigning ADA information is missing less than 50% of the time (South Boston, East Boston, and West Roxbury), reporting a 47 pp decrease in the probability that a defendant receives a subsequent criminal complaint within two years within this sample ($p < .05$). Column (3) of Panel B replicates this analysis within years where arraigning ADA information is missing less than 60% of the time (2004, 2006-2008), reporting a 30 pp decrease in the probability that a defendant receives a subsequent criminal complaint within two years (not significant).

Finally, Panel C progressively expands the sample by imputing arraigning ADA assignment using the strategies detailed in Appendix C.2. In our largest imputed sample ("Imputation 4," containing 147,080 observations or 72% of nonviolent misdemeanor cases meeting all other sample criteria), 24% of prosecuted cases are missing arraigning ADA information, and 22% of not prosecuted cases are missing arraigning ADA information. Second stage estimates for the imputation samples range between -0.27 and -0.31, with p < .05. Overall, the results reported in Table 6 suggest that our primary estimates are not being driven by selection bias from missing arraigning ADA information.[16]

5 Mechanisms

We consider multiple mechanisms that may be driving our findings. First, nonprosecution reduces the probability that a defendant will receive a DCJIS criminal record of their misdemeanor charge. Cases dismissed prior to formal arraignment do not receive DCJIS records in Massachusetts; cases that proceed to formal arraignment do receive DCJIS records. In our data we can only identify day of final disposition, not whether a final disposition that occurs on the day of arraignment occurs before or after formal arraignment. We expect, however, that cases not prosecuted under our definition will have significantly lower rates of DCJIS record acquisition, relative to cases that are prosecuted. As reported in Table 1, 37% of cases that are not prosecuted have a DCJIS record, relative to 77% of cases that are prosecuted.[17] This is also true for marginal defendants. As reported in Appendix Table B.4, Panel B, a ten percentage point increase in ADA leniency results in a 2.9 percentage point decrease (42%) in the probability that a defendant receives a misdemeanor charge record in the DCJIS database. Prosecuted defendants are likely aware that they have acquired a criminal record at arraignment, and that this record could decrease their employment prospects, potentially decreasing their incentives to refrain from engaging in criminal activity (Smith and Broege 2020; Herring and Smith 2022). Nonprosecuted defendants are likely aware that they have avoided acquiring a criminal record of charges, and that a clean record increases their em-

[16] In Appendix Section C.2 we also show that if we extrapolate from the first stage and reduced form estimates in the main sample, we can exactly re-create the differences in the probability of a criminal complaint within two years in the sample missing ADA information. We believe this further suggests that ADA missingness is not biasing our results. We thank an anonymous referee for this suggestion.

[17] Although all prosecuted cases should have DCJIS records, SCDAO and DCJIS records were matched on docket numbers, and there appears to be considerable human error in docket number entry. However, we have no reason to expect systematic bias in this data entry error.

ployment prospects, potentially increasing the salience of the returns to desistance. These behavioral incentives are likely to be strongest in the case of defendants (both prosecuted and nonprosecuted) without prior criminal records.

Second, nonprosecution eliminates the possibility that defendants will spend a lengthy period of time in the criminal justice system with an open case. As reported in Table 1, the prosecuted cases in our sample take on average 185 days to resolve and have on average four case events; nonprosecuted cases, mechanically, take 0 days to resolve and have only 1 case event. This is again also true for marginal defendants. As reported in Appendix Table B.4, Panel B, a 10 percentage point increase in ADA leniency results in 8.5 fewer days to disposition (50%) and 0.17 fewer case events (58%). Time spent in the criminal justice system—attending hearings and meetings with lawyers, for example—may disrupt defendants' work and family lives, increasing the risk of reoffending.

Third, nonprosecution eliminates the possibility that a defendant will receive a conviction in their case. As reported in Table 1, 26% of prosecuted cases in our sample result in a conviction; 0% of nonprosecuted cases result in conviction. As reported in Appendix Table B.4, Panel B, a 10 percentage point increase in ADA leniency results in a 1.5 percentage point decrease in the probability of conviction (72%) for marginal defendants. Criminal records of misdemeanor convictions may further damage defendants' labor market prospects beyond records of criminal charges, additionally raising the risk of reoffending. However, sentences of incarceration and probation, conditional on conviction, are not likely driving the impacts we observe. We were able to secure sentencing data from the Massachusetts trial courts for SCDAO cases initiated between the years 2015-2019. In our main two-year estimation sample, among prosecuted nonviolent misdemeanor defendants with arraignment dates during that period, 12.3% received convictions. Among those who received convictions, 22.7% (only 2.8% of all prosecuted defendants) received sentences of probation, and 27.2% (only 3.3% of all prosecuted defendants) received sentences of incarceration.

We attempt to distinguish between these mechanisms in Table 4 by subsetting defendants on the basis of their criminal histories. If the disruption caused by an open case were driving the effects we observe, then we would expect to see large negative effects of nonprosecution across all categories of defendants, regardless of their criminal histories. However, this is not the case; the point estimates of the effect of nonprosecution for those with prior criminal histories are negative but small or suggestively positive (columns (2), (4), and (5)).

If misdemeanor conviction records were driving the effects we observe, then we would expect to see large negative effects for defendants acquiring their first conviction record, with less impact on those who already have prior convictions. However, as reported in columns (4) and (5), we do not have strong evidence for this mechanism; among the set of defendants with prior criminal records, 2SLS coefficients are positive although statistically insignificant for defendants with and without prior conviction records.

Finally, if the acquisition of criminal records of charges were driving the effects we observe, then we would expect to see substantial negative effects for defendants acquiring their first criminal record of any kind, with less impact on those who already have criminal histories. The estimates in Table 4 indicate that this is the case: nonprosecution has a large and weakly significant negative effect on subsequent criminal activity for defendants who have no prior criminal complaint in Suffolk County (column (1)). Nonprosecution has an even larger and more precisely estimated negative effect on subsequent criminal activity for defendants who have no prior DCJIS record (column (2)). Although the confidence intervals on the point estimates overlap for defendants with and without criminal histories, it is important to note that those with previous criminal histories also have higher base rates of subsequent complaints within two years. Estimated reductions in criminal activity for first-time defendants thus represent larger proportional changes (around 80% over control complier means), suggestive of an economically meaningful effect of the acquisition of criminal records of misdemeanor charges on reoffending. This is consistent with previous evidence that even misdemeanors that do not lead to convictions can adversely impact employer callback behavior and defendant job search behavior (Uggen et al. 2014; Smith and Broege

2020).

The larger proportional effect for first-time defendants reported in Table 4 also helps to explain why the magnitude of our estimated effects is comparable to the effect magnitude of avoiding a felony conviction for first-time felony defendants (Mueller-Smith and Schnepel, 2021). The Mueller-Smith and Schnepel (2021) sample excluded defendants with prior felony charges or convictions, but did not exclude defendants with prior misdemeanor charges or convictions, potentially dampening the effects of avoiding a felony conviction on subsequent reoffending.

Although our data do not allow us to fully rule out either the disruption caused by an open case or misdemeanor conviction records as causal mechanisms, these do not appear to be the main drivers of our observed recidivism effects. While it is possible that there is simply less scope for these mechanisms to work among those with prior criminal histories, we also cannot provide evidence for their existence. There are also other mechanisms that our data do not allow us to eliminate. In particular, it is possible that nonprosecuted defendants are more likely to move out of Suffolk County (or choose to pursue criminal offending outside of Suffolk County); our data do not allow us to rule out this mechanism. We do not, however, have any reason to think that this would be the case.

5.1 Ratcheting

Does the reduction in subsequent criminal justice contact we observe result from changes in defendant criminal behavior or purely from a "ratcheting" effect, i.e., a law enforcement reaction to individuals with knowable criminal histories? In the former case, nonviolent misdemeanor defendants who are not prosecuted commit fewer subsequent offenses, relative to defendants who are prosecuted, perhaps because of behavioral changes induced by criminal record acquisition. In the latter case, nonviolent misdemeanor defendants who are not prosecuted are equally likely to commit subsequent offenses, relative to defendants who are prosecuted, but police officers are more likely to bring new criminal complaints against formerly prosecuted defendants, who acquired criminal records of charges observable to officers. In this latter case, there are clear benefits to defendants from nonprosecution, but the public safety benefits are less clear.

This question is difficult to answer without access to information on a) defendants' "true" criminal activity and b) law enforcement encounters with defendants in our sample that did not lead to criminal complaints. However, we can offer some suggestive findings. First, we leverage the distinction between discretionary and non-discretionary arrests. Police officers arguably have greater discretion to make arrests in some situations (e.g., no victims/witnesses, no 911 call, less serious offense) than in others (e.g., victims/witnesses, 911 call, more serious offense). If the downstream effects we observe were driven largely by police officers' reactions to defendants' prior criminal histories, we would expect to see large downstream effects in situations wherein the police have more discretion, and less impact in situations wherein they have less discretion. To define discretionary and non-discretionary arrests, we use the machine learning-based classifications reported in Abdul-Razzak and Hallberg (2021), predicting proportions of discretionary and non-discretionary arrests across crime types based on arrest and stop reports. We categorize a criminal complaint as "non-discretionary" if any offense is that complaint is classified as non-discretionary using the classifications in Abdul-Razzak and Hallberg (2021); we categorize a criminal complaint as "discretionary" if all offenses in the complaint were classifiable and no offenses were classified as non-discretionary. We then repeat our main 2SLS analysis of the impact of nonprosecution on the probability of a new criminal complaint within two years separately for discretionary and non-discretionary complaints. In Table 5 Panel A, columns (1) and (2), we see that the nonprosecution of marginal first-time defendants results in a very large, statistically significant decrease in new non-discretionary criminal complaints, and a suggestively negative but not statistically significant decrease for new discretionary criminal complaints. We see no similar pattern for repeat offenders, who are less likely to benefit from nonprosecution.

We repeat this exercise for the categories of victim and victimless offenses defined earlier in Table 1. Police officers arguably have less discretion to issue criminal complaints for offenses with victims, relative to offenses without victims. This categorization partially but not completely overlaps the discretionary/non-discretionary categorization defined above. As reported in Table 5 Panel A, columns (3) and (4), we see a very similar story: large statistically significant decreases in the probability of a new complaint within two years for offenses with victims, but smaller and insignificant decreases for offenses without victims. We again see no similar distinction for repeat offenders.

Taken collectively, these findings are not consistent with the hypothesis that the downstream effects we estimate are being driven only by police officers' reactions to defendants' observable criminal histories. We see large, statistically significant downstream effects in situations wherein the police have less discretion to exercise leniency in response to a defendant's lack of an observable criminal history; we see less impact in situations wherein the police have less discretion, although we cannot rule out similar effects. This implies that our findings are more likely explained by changes in defendants' criminal behavior, although we do not completely rule out a role for "ratcheting."

Whether the downstream effects we observe could be driven largely by the police reaction to defendants' observable criminal histories also depends crucially on what officers can observe about those histories. We interviewed individuals who work or had worked for the Boston Police Department and the Cambridge Police Department to get an understanding of the information available in patrol cars' Mobile Data Terminals (MDTs). Both representatives confirmed that Massachusetts police officers have access to state DCJIS records in patrol cars. They both also confirmed that officers also have access to electronic records of criminal complaints made by their own departments. This implies there will be an informational difference when an officer encounters a defendant whose prior complaint was brought by the officer's own agency, relative to a different agency. In the former case, the officer will have access to the defendant's prior complaint even if that complaint was not prosecuted (and therefore was not recorded in the statewide DCJIS database). In the latter case, the officer would likely only have access to the prior complaint if it was prosecuted.

If the downstream effects that we observe were due largely to police officers' responses to observable prior complaints (and defendants' choices of locations for criminal activity were unresponsive to agency boundaries), then we would expect large negative effects when officers do not have information about previous complaints that were not prosecuted (i.e., when officers are employed by agencies that are different from the agency that brought the original complaint), and no impact when officers do have information about previous complaints. Table 5 Panel A, columns (5) and (6) explore this distinction. The point estimates indicate an 83% decrease in new criminal complaints for same agency complaints, relative to compiler means, and a 76% decrease for different agency complaints, although the estimates are noisy. Only the different police agency coefficient is statistically significant, although using conventional cluster-robust standard errors, the Anderson-Rubin confidence interval for this coefficient includes 0. We again see no similar distinction for repeat offenders. These estimates are not supportive of the hypothesis that the downstream effects that we observe are due only to police officers' responses to observable prior complaints.[18]

In short, while we cannot rule out "ratcheting" behavior by police officers in response to defendants' observable criminal histories, the analyses reported here are not consistent with the hypothesis that our findings are only or even primarily driven by ratcheting behavior. Misdemeanor prosecution does appear to lead to behavioral changes by marginal first-time nonviolent misdemeanor defendants.

6 Policy Relevance and Moving beyond the LATE

Our 2SLS estimates give us a weighted average of the effect of nonprosecution among those defendants induced into nonprosecution by being (as-if randomly) assigned a more lenient ADA at arraignment. The decision to prosecute or not prosecute a defendant against whom

[18]Although in theory it is possible that clerk magistrates respond to defendants' criminal histories in their reviews of police officers' applications for criminal complaints, our conversations with SCDAO staff indicate that this is highly unlikely.

a criminal complaint has been issued is a decision that rests squarely with the office of the district attorney in that jurisdiction. Conditional on the set of behaviors that are considered criminal, and the behavior of police in arresting individuals suspected of committing those crimes, the only policy lever available to change nonprosecution rates is to increase the leniency of the individuals within a district attorney's office who make the prosecution decision. This LATE estimate is thus also a policy-relevant treatment effect (PRTE) (Heckman and Vytlacil 2001; Heckman and Urzua 2010; Cornelissen et al. 2016).

As we increase leniency, we would presumably be drawing different marginal defendants into nonprosecution. Our LATE estimates do not tell us directly how these marginal defendants may differ in their treatment effects from defendants likely to be on the margin of prosecution for less lenient ADAs, or what would happen with a large increase in average leniency. We explore this question in two ways. First, we estimate marginal treatment effects (MTEs). These MTEs give insight into what might happen if we implemented a policy that increased ADA leniency. The MTE estimates rely on stronger assumptions than our 2SLS estimates. We also cannot extrapolate beyond the data that we have: the MTEs are only estimated for predicted probabilities of nonprosecution for which we see both prosecuted and nonprosecuted individuals—the common support of the propensity score for nonprosecution.

Second, we consider the effects of a policy change. Several district attorney's offices around the country have begun to implement policies of presumptive nonprosecution for certain (usually nonviolent misdemeanor) offenses. To the extent that such policies still allow room for ADA discretion, the presumption of nonprosecution may be applied largely to marginal nonviolent misdemeanor defendants, and may have similar effects as those estimated above. However, to the extent that the policies expand the set of marginal defendants beyond those in our sample, and/or are applied to non-marginal defendants, the policies may have different effects. During her 2018 election campaign, Rachael Rollins campaigned on a platform that included a presumption of nonprosecution for nonviolent misdemeanor offenses. We use her inauguration on January 2, 2019 as District Attorney of Suffolk County as a natural experiment to explore such policy effects and understand the impacts of increases in nonprosecution.

6.1 Marginal Treatment Effects

Because defendants are (as-if) randomly assigned to a large number of ADAs with different leniency rates, we can trace out the effects of nonprosecution along different margins of the unobserved resistance to nonprosecution by estimating marginal treatment effects—the derivative of the probability of a criminal complaint within two years with respect to the predicted probability of nonprosecution (Heckman and Vytlacil 2005; Heckman et al. 2006; see Appendix A.4 for more details on the derivation of the MTE in the potential outcomes framework).[19] The MTEs thus show how subsequent criminal justice contact varies across defendants who are induced into nonprosecution as the predicted probability of nonprosecution varies with the instrument. At higher levels of the unobserved resistance to nonprosecution, we estimate effects for defendants on the margin for only the most-lenient ADAs (i.e., defendants closer to the never-takers who are prosecuted most often), giving us an idea of what would happen if we expanded leniency towards this group.

Figure 4a shows the support of the predicted probability of nonprosecution (the propensity score) for prosecuted and nonprosecuted defendants. We can only trace out the MTEs along this range of common support.[20] There are many potential functional forms one could assume for the empirical MTE specification. Our main estimation is based on a cubic polynomial specification, although in Appendix Figure B.3 we show how the MTE estimates vary with other assumptions for the functional form. Estimating and interpreting MTEs also requires a strict monotonicity assumption and additive separability between observed and unobserved heterogeneity in the treatment effects, stronger assumptions than required

[19] In practice, we use the Stata package `mtefe` (Andresen, 2019). See Doyle Jr (2007), Maestas et al. (2013), French and Song (2014), Arnold et al. (2018), and Bhuller et al. (2020) for empirical examples of MTE estimation in leniency designs.

[20] The estimates can become imprecise at the extreme ends of this distribution given smaller numbers of ADAs, so when estimating the MTEs we trim the top and bottom one percentiles of this common support distribution.

to interpret our 2SLS estimates (Brinch et al. (2017)). Using the test of Frandsen et al. (2019) we could not reject the null of strict monotonicity holding in 6 out of 9 courts. Our main MTE analysis is estimated in our full data; we also repeat the analysis restricted to the 6 courts where we could not reject this null (see Appendix Table B.3) and find very similar results. These assumptions are quite stringent, however, and thus we see these MTE estimates as interesting but only suggestive.

Figure 4b shows the estimated MTEs with the cubic polynomical specification. The MTEs appear to decline monotonically as the unobserved resistance to nonprosecution increases—marginal defendants who are closer to never-takers in the IV framework experience, if anything, larger decreases in recidivism when not prosecuted. Appendix Figure B.3 shows how the shape of the estimated MTE varies with different functional forms. MTE estimates are sensitive to specification but we never see an upward sloping MTE; estimates are either monotonically declining or flat, with all estimates indicating that there is still a reduction in subsequent criminal involvement even at relatively high levels of the unobserved resistance to nonprosecution. These suggestive estimates imply that increasing the leniency of ADA nonprosecution decisions would not cause increases in recidivism and, if anything, might cause even larger decreases in subsequent criminal justice contact than the average estimates reported earlier. We can also express other treatment effect parameters as weighted averages of the MTEs, such as the (overall) average treatment effect (ATE), average treatment on the treated (ATT), and average treatment on the untreated (ATUT). We rescale the weights so that they integrate to one over the common support region shown in Figure 4a and estimate these three treatment effects (Carneiro et al., 2011; Andresen, 2019). We report the estimates in the upper right corner of Figure 4b, along with the rescaled LATE estimate.

The suggestively downward sloping MTE models for some specifications, and the fact that our 2SLS estimates were larger than our OLS estimates, lead to a question of prosecutors' objective function. In the case of bail decisions, magistrates and judges are supposed to be focused on reducing failure to appear and/or pretrial misconduct (Kleinberg et al. 2018). In the decision about whether to prosecute a defendant, the objective function is less clear. For example, prosecutors might care about deterrence (reducing future recidivism), the probability of conviction, a defendant's culpability, and/or other potential considerations. If prosecutors were only trying to deter future criminal behavior, then the negative selection and downward sloping MTEs would imply that they are not doing so most efficiently. If prosecutors were only trying to maximize conviction rates, then we would expect that individuals prosecuted by more lenient ADAs would be more likely to be convicted, but we saw in Appendix Table B.4 no correlation between ADA leniency and conviction among prosecuted defendants. Alternatively, (some) ADAs may have incorrect priors about the probability of conviction, although we do not see any change in our estimates when we construct ADA leniency by experience level (Appendix Table B.9). Instead, our results imply that there are other factors that enter into prosecutors' objective function, like unobserved culpability, that explain both the downward sloping MTEs and the magnitude difference between IV and OLS estimates—those defendants that prosecutors see as most deserving of a second chance may also be most likely to recidivate.

6.2 Effects of a Presumption of Nonprosecution

In this subsection we explore the impacts of the inauguration of Rachael Rollins as District Attorney of Suffolk County on January 2, 2019. During her 2018 election campaign, then-candidate Rollins campaigned on a platform of presumptive nonprosecution for nonviolent misdemeanor offenses. After taking office Rollins remained publicly committed to this policy, leading some ADAs to resign from the office and others to be hired as replacements. On March 25, 2019, the office issued a public memo announcing a new policy of presumptive nonprosecution for a set of nonviolent misdemeanor offenses.

Figure 5 reports monthly changes in nonprosecution rates, relative to the month of December 2018, for cases involving nonviolent misdemeanor and nonviolent felony complaints initiated between January 1, 2018 and September 1, 2019.

Nonprosecution rates for both nonviolent misdemeanors and nonviolent felonies were relatively flat during the year prior to the arrival of District Attorney Rollins; the estimated linear time trends for the pretreatment coefficients are -0.00005 for nonviolent misdemeanors (se = .002) and -0.00094 for nonviolent felonies (se = .001). The event study plots and estimated trends in the pretreatment coefficients suggest that nonprosecution rates for nonviolent misdemeanors were not already trending upward, and were paralleling nonprosecution rates for nonviolent felonies, prior to Rollins' inauguration. Appendix Figures B.5 and B.6 show that the number of SCDAO cases involving nonviolent misdemeanor and nonviolent felony complaints and the number of arrests made by the Boston Police Department for nonviolent offenses were steady during this period.

After Rollins' inauguration, nonprosecution rates for nonviolent misdemeanor complaints began to climb steeply, peaking in May 2019; the same is not true for nonviolent felonies, which were not included in the campaign platform's presumption of nonprosecution.[21] We leverage the sharp post-Rollins increase in nonprosecution rates for nonviolent misdemeanor complaints to estimate the effects of presumptive nonprosecution on subsequent criminal complaints within one year.[22]

In Table 7 column (1), we first report OLS estimates of the effect of nonprosecution on one-year complaint rates for nonviolent misdemeanor defendants whose complaints were initiated between January 1, 2018 and September 1, 2019, finding an 8 pp reduction in subsequent complaint rates (27% reduction relative to the 30% one-year complaint rate for prosecuted defendants; $p < .01$). In column (2) we then estimate a 2SLS model using the

[21] As reported in Appendix Figure B.7, nonprosecution rates rose after Rollins' inauguration both for the set of nonviolent misdemeanor offenses later included in the March 2019 policy memo and for all other nonviolent misdemeanor offenses, perhaps because of ADA turnover and Rollins' public commitment to presumptive nonprosecution for nonviolent misdemeanor offenses. Appendix Figure B.7 also reveals that nonprosecution rates for nonviolent misdemeanors not included in the March policy memo declined less steeply after May 2019, relative to those for nonviolent misdemeanors included in the memo, suggesting that emerging law enforcement criticism of the March policy memo may have caused some reductions in nonprosecution for the offenses highlighted in that memo.

[22] A one-year post-arraignment window is used due to both limited data post-inauguration and the COVID-19 pandemic.

Rollins inauguration as an instrument for the nonprosecution of nonviolent misdemeanor complaints after December 2018. In the first stage (Panel B), we find an average 6 pp post-Rollins increase in the nonprosecution rate for nonviolent misdemeanor complaints (17% increase relative to the 36% pre-Rollins nonprosecution rate for these cases; $p < .01$), with an F-statistic of 58.44. In the second stage (Panel A), we find a 59 pp reduction in one-year complaint rates for those marginal nonviolent misdemeanor defendants whose cases were not prosecuted after the Rollins inauguration ($p < .01$). In column (3) we use nonviolent felony complaints as a control group for nonviolent misdemeanor complaints, estimating an IV difference-in-differences model using the Rollins inauguration as an instrument for the nonprosecution of nonviolent misdemeanor complaints after December 2018, relative to the nonprosecution of nonviolent felonies. In the first stage (Panel B), we find an average 5 pp post-Rollins increase in the nonprosecution rate for nonviolent misdemeanor complaints ($p < .01$), relative to the 1 pp post-Rollins increase in the nonprosecution rate for nonviolent felony complaints (not significant), with an F-statistic of 28.57. In the second stage (Panel A), we find a 52 pp reduction in one-year rearrest rates for those marginal nonviolent misdemeanor defendants whose cases were not prosecuted after the Rollins inauguration ($p < .10$), relative to the essentially unchanged post-Rollins one-year rearrest rate for nonviolent felony complaints.

Columns (4) and (5) report reduced form models of the effect of the Rollins inauguration on one-year rearrest rates for nonviolent misdemeanor defendants, with and without nonviolent felony complaints as a control group. In column (4), we see a 3 pp post-Rollins reduction in one-year rearrest rates for nonviolent misdemeanor defendants (12% relative to the 26% one-year rearrest rate for pre-Rollins nonviolent misdemeanor defendants, $p < .01$). In column (5), we again see a 3 pp post-Rollins reduction in one-year rearrest rates for nonviolent misdemeanor defendants ($p < .10$), relative to essentially unchanged one-year rearrest rates for nonviolent felony defendants.

Similar to our main estimates of the impacts of increased ADA leniency at arraignment, these estimates suggest that policies introducing a presumption of nonprosecution for nonviolent misdemeanor offenses may have social benefits. The increases in nonprosecution of nonviolent misdemeanor offenses induced by the Rollins inauguration appear to have decreased the rates at which defendants were issued new criminal complaints within one year of the current case.

It is also possible, however, that a policy change to reduce the prosecution of nonviolent misdemeanors could increase the number of crimes committed by other individuals who are not in our data by reducing general deterrence. Appendix Figure B.8 shows the effects of District Attorney Rollins' inauguration on crimes reported by the Boston Police Department. We focus on the types of offenses for which the expected probability of prosecution might have decreased after the Rollins inauguration. The data include crime reports from January 2017 through February 2020 (before COVID-19). We group incidents into the following categories: property damage, theft and fraud, disorder, drug, and other offenses. Overall, we find significant reductions in reports of property damage and reports of theft/fraud. There is no evidence of an increase in any of these crime types.

Overall, we interpret these effects of District Attorney Rollins' inauguration and implementation of policies that reduced the prosecution of nonviolent misdemeanors as suggestive evidence that this policy shift, a relatively large expansion in leniency, reduced the subsequent average criminal justice involvement of the broader pool of defendants now experiencing leniency. Effects on reported crime are noisy, but there is no evidence that this policy change had detrimental effects on public safety. It will be important to track changes over time in this setting and elsewhere, to more fully understand what trade-offs, if any, exist.

7 Discussion

Misdemeanor cases make up over 80 percent of the cases processed by the U.S. criminal justice system. Yet we know little about the causal impacts of misdemeanor prosecution or nonprosecution. We report the first estimates of the causal effects of misdemeanor nonprosecution on rates and numbers of post-arraignment criminal complaints. To do this, we leverage the as-if random assignment of nonviolent misdemeanor cases to arraigning ADAs in a large urban district attorney's office. Our findings imply that not prosecuting marginal nonviolent misdemeanor defendants substantially reduces their subsequent criminal justice contact, or, in other words, that prosecuting marginal nonviolent misdemeanor defendants substantially increases their subsequent criminal justice contact.

The key policy question that motivated this study is whether scaling back the prosecution of nonviolent misdemeanor prosecution would enhance or reduce public safety. Our findings indicate that the observed increases in criminal justice contact after the prosecution of marginal nonviolent misdemeanor defendants are due at least in part to changes in defendants' behavior after the acquisition of their first criminal record of charges. These findings are troubling, given the volume of misdemeanor prosecutions pursued in the United States. We may in fact be undermining public safety by criminalizing relatively minor forms of misbehavior.

Our results suggest that inducing arraigning ADAs to be more lenient in their prosecution decisions could yield net social benefits. Preliminary evidence on the effects of a related policy change in Suffolk County—a presumption of nonprosecution for nonviolent misdemeanor offenses—supports this policy implication. We look forward to seeing future work on the longer-run effects of the SCDAO policy, and on the effects of similar prosecutor-led reforms in other contexts.

References

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

Abdul-Razzak, N. and K. Hallberg (2021). Unpacking the promise and limitations of behavioral health interventions on interactions with law enforcement. Technical report, Working Paper.

Agan, A., M. Freedman, and E. Owens (2021). Is your lawyer a lemon? incentives and selection in the public provision of criminal defense. *Review of Economics and Statistics 103*(2), 294–309.

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

Ahrens, A., C. Hansen, and M. Schaffer (2019, January). `ivlasso and pdslasso`: Programs for post-selection and post-regularization OLS and IV estimation and inference.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical statistics 20*(1), 46–63.

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

Andresen, M. E. (2019). `MTEFE`: Stata module to compute marginal treatment effects with factor variables. This version November 2020.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*, 727–753.

Angrist, J. and B. Frandsen (2020). Machine labor. Working Paper 26584, National Bureau of Economic Research.

Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association 90*(430), 431–442.

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

Angrist, J. D. and A. B. Keueger (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics 133*(4), 1885–1932.

Arnold, D., W. S. Dobbie, and P. Hull (2020). Measuring racial discrimination in bail decisions. Technical report, National Bureau of Economic Research.

Augustine, E., J. Lacoe, S. Raphael, and A. Skog (2021). The impact of felony diversion in san francisco. Technical report, California Policy Lab Working Paper.

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

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

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.

Bhuller, M., G. Dahl, K. Loken, and M. Mogstad (2020). Incarceration, recidivism and employment. *Journal of Political Economy 128*(4).

Bishop, E., B. Hopkins, C. Obiofuma, and F. Owusu (2020, September). Racial dispariies in the massachusetts criminal system. Technical report, Criminal Justice Policy Program Harvard Law School.

Björklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics 69*, 42–49.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022). When is tsls actually late? Working Paper 29709.

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

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

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

Chalfin, A. and J. McCrary (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature 55*(1), 5–48.

Chao, J. C. and N. R. Swanson (2005). Consistent estimation with a large number of weak instruments. *Econometrica 73*(5), 1673–1692.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review 104*(9), 2593–2632.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics 41*, 47–60.

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

Davidson, R. and J. G. MacKinnon (2014). Confidence sets based on inverting anderson–rubin tests. *The Econometrics Journal 17*(2), S39–S58.

de Chaisemartin, C. (2017). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics 8*(2), 367–396.

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

Doleac, J. L. (2020). Encouraging desistance from crime. Working paper.

Doyle Jr, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review 97*(5), 1583–1610.

Dusek, L. and C. Traxler (2020). Learning from law enforcement.

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

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

French, E. and J. Song (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy 6*(2), 291–337.

Gehrsitz, M. (2017). Speeding, punishment, and recidivism: Evidence from a regression discontinuity design. *The Journal of Law and Economics 60*(3), 497–528.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2022). Contamination bias in linear regressions. Working Paper 30108.

Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *American Economic Review 105*(4), 1581–1617.

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

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

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *American Economic Review 91*(2), 107–111.

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

Herring, C. and S. S. Smith (2022). The limits of ban-the-box legislation.

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

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

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018). Human decisions and machine predictions. *The quarterly journal of economics 133*(1), 237–293.

Kohler-Hausmann, I. (2013). Misdemeanor justice: Control without conviction. *American Journal of Sociology 119*(2), 351–393.

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

Landersø, R., H. S. Nielsen, and M. Simonsen (2017). School starting age and the crime-age profile. *The Economic Journal 127*(602), 1096–1118.

Laub, J. H. and R. J. Sampson (2001). Understanding desistance from crime. *Crime and justice 28*, 1–69.

Leasure, P. (2019). Misdemeanor records and employment outcomes: An experimental study. *Crime & Delinquency 65*(13), 1850–1872.

Lee, D. S., M. J. Moreira, J. McCrary, and J. Porter (2020, October). *arXiv e-prints*. arXiv:2010.05058.

Maestas, N., K. J. Mullen, and A. Strand (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review 103*(5), 1797–1829.

Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics 10*, 577–613.

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

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association 78*(381), 47–55.

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

Nagin, D. S., F. T. Cullen, and C. L. Jonson (2009). Imprisonment and reoffending. *Crime and Justice 38*(1), 115–200.

Natapoff, A. (2018). *Punishment Without Crime: How Our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal*. New York: Basic Books.

Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology 108*(5), 937–975.

Pager, D. (2008). *Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration*. University of Chicago Press.

Raphael, S. and M. A. Stoll (2014). A new approach to reducing incarceration while maintaining low rates of crime. Hamilton Project Discussion Paper 2014-03.

Rehavi, M. M. and S. B. Starr (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy 122*(6).

Rempel, M., M. Labriola, P. Hunt, R. C. Davis, W. A. Reich, and S. Cherney (2018). *NIJ's Multisite Evaluation of Prosecutor-led Diversion Programs: Strategies, Impacts, and Cost-effectiveness*. Center for Court Innovation New York, NY.

Rivera, R. (2021). A more diverse police academy improves job performance: The effect of minority peers on future arrest quantity and quality. Technical report, Working Paper, Columbia University, New York.

Sloan, C. (2020a). How much does your prosecutor matter? an estimate of prosecutorial discretion. Working paper.

Sloan, C. (2020b). Racial bias by prosecutors: Evidence from random assignment. Working paper.

Smith, S. S. and N. C. Broege (2020). Searching for work with a criminal record. *Social Problems 67*(2), 208–232.

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

Tuttle, C. (2021). Racial disparities in federal sentencing: Evidence from drug mandatory minimums. Working paper.

Uggen, C., M. Vuolo, S. Lageson, E. Ruhland, and H. K. Whitham (2014). The edge of stigma: An experimental audit of the effects of low-level criminal records on employment. *Criminology 52*(4), 627–654.

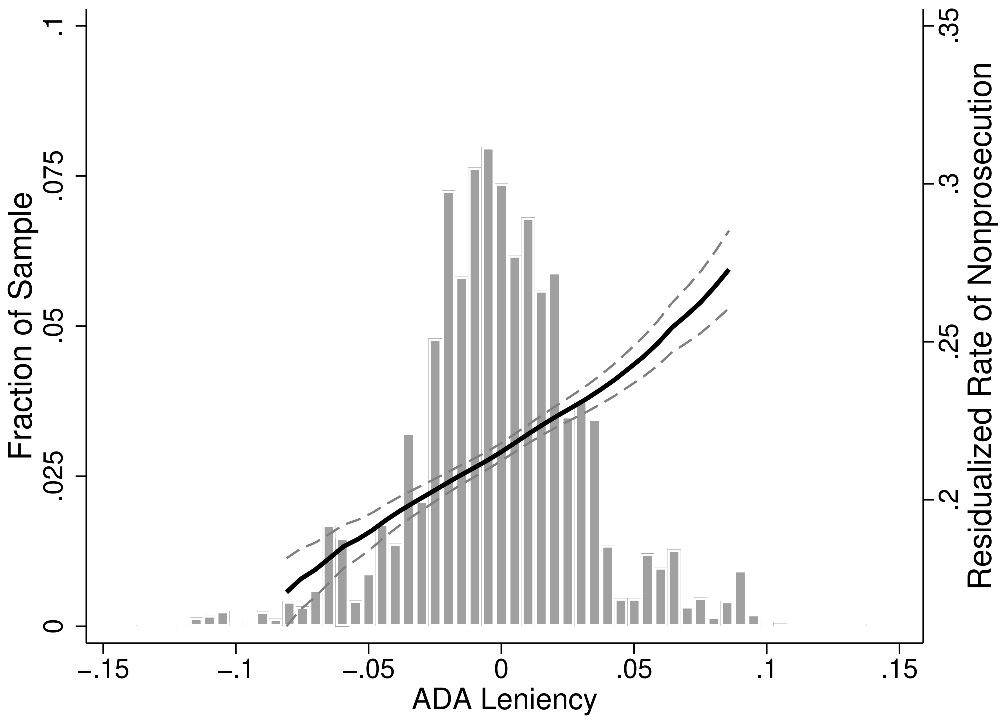


Figure 1: ADA Leniency and Nonprosecution

Notes. This figure shows the distribution of our leave-out mean measure of ADA nonproseuction ("leniency"), residualized by court-by-year-month and court-by-day-of-week. More lenient ADAs have higher rates of not prosecuting nonviolent misdemeanor cases. The solid line is a local linear regression of nonprosecution on ADA leniency, along with the 95% confidence interval, estimated from the 1st to 99th percentiles of ADA leniency—a local linear version of our first stage. A case assigned to a more lenient ADA (computed using all cases except the current case and other cases with the same defendant) has a higher likelihood of being not prosecuted.

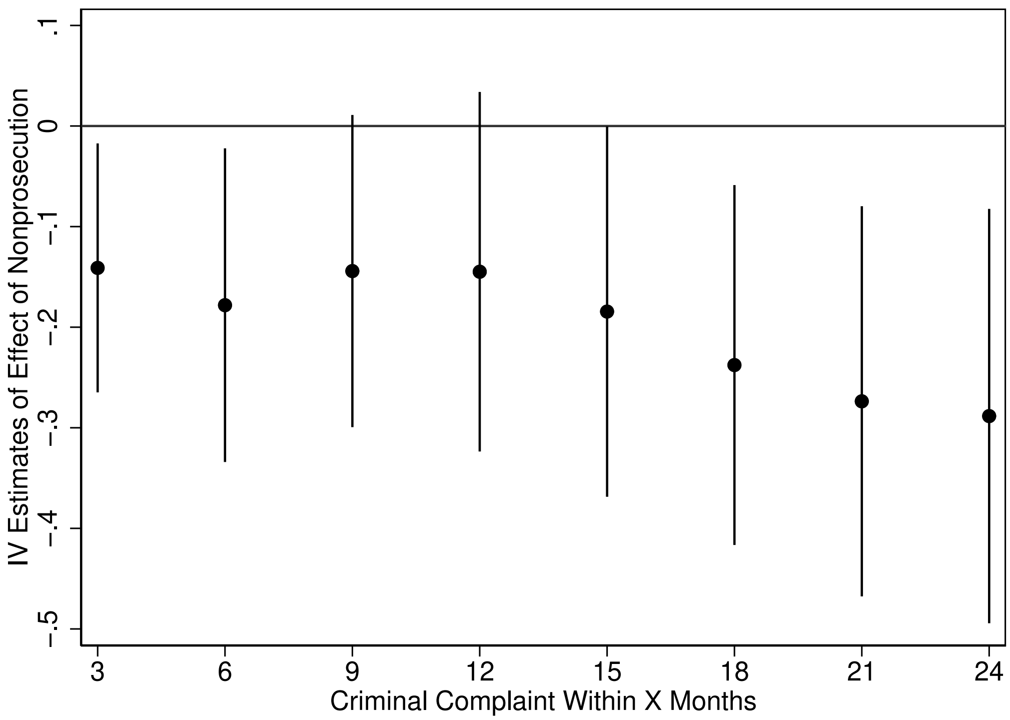


Figure 2: LATE by Months

Notes. This figure shows the local average treatment effect of nonprosecution on the likelihood of a new criminal complaint (y-axis) within a given number of months after the initial arraignment date (x-axis). Estimates are based on 2SLS regressions including covariates (the equivalent of column (4) in Table 3); the estimation sample is the same as in Table 3. Dots report coefficients; lines report 95% confidence intervals.

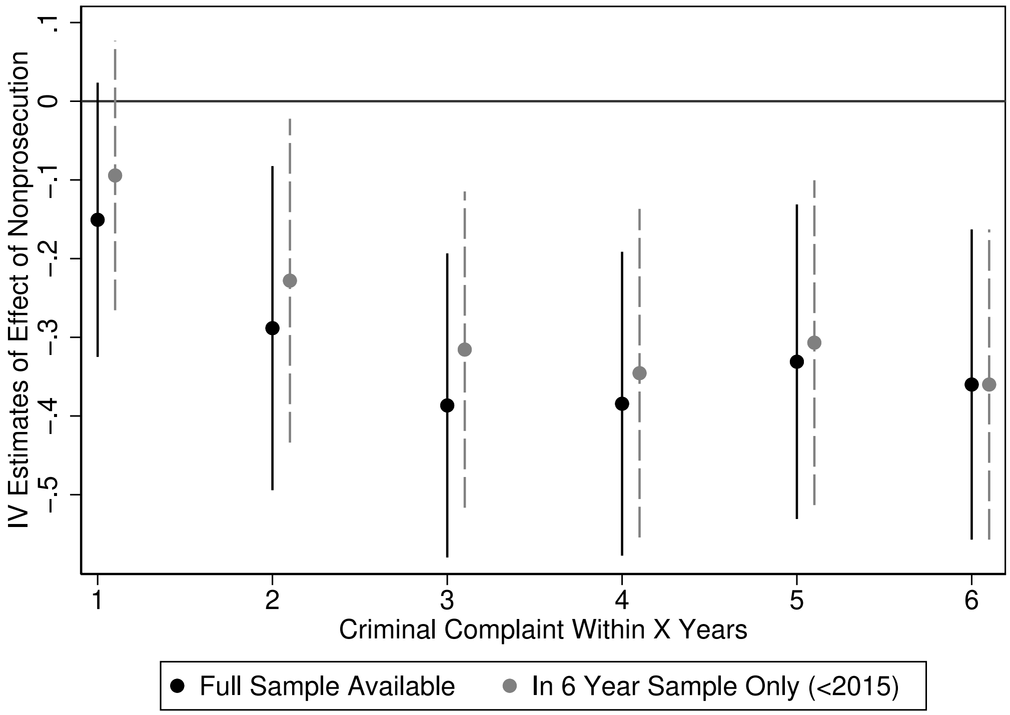
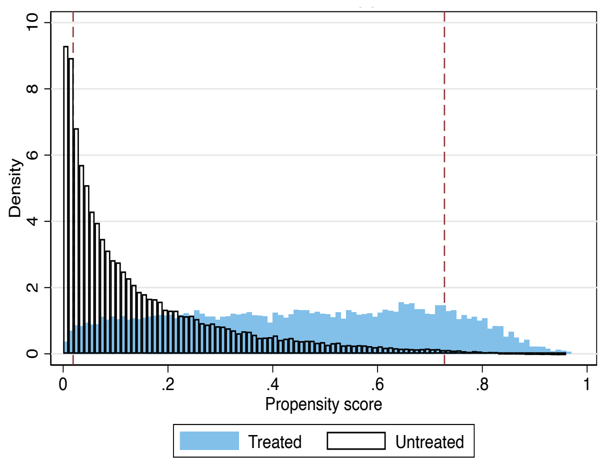
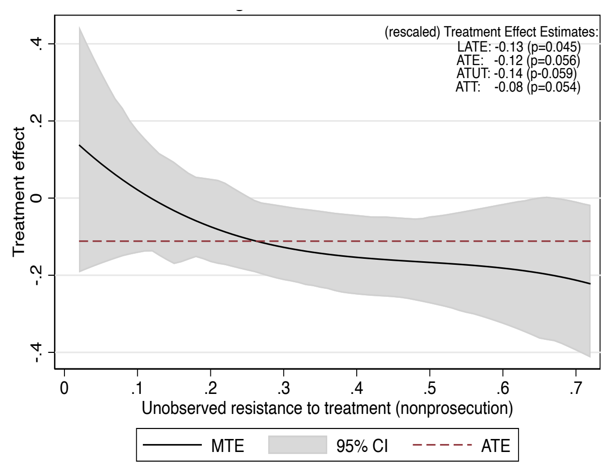


Figure 3: LATE With Different Time Horizons

Notes. This figure shows the local average treatment effect of nonprosecution on the likelihood of a new criminal complaint (y-axis) within a given number of years after the initial arraignment date (x-axis). Estimates are based on 2SLS regressions including covariates (the equivalent of column (4) in Table 3). Dots report coefficients; lines report 95% confidence intervals. Coefficients represented using darker dots are estimated using different samples, allowing for larger samples of defendants for shorter time horizons. Coefficients represented using lighter dots restrict the sample to defendants with 6 years of post-arraignment data.



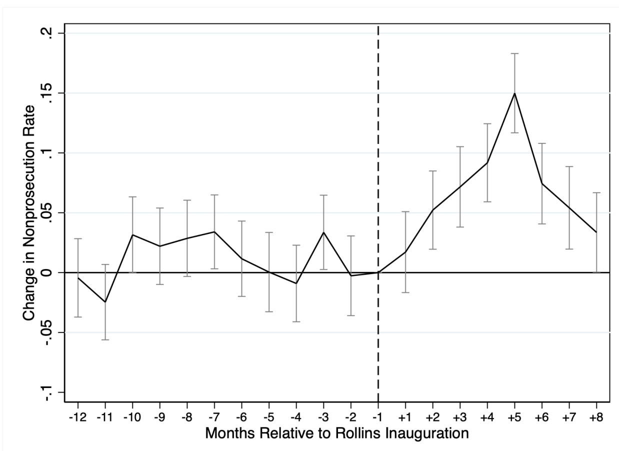
(a) Common Support



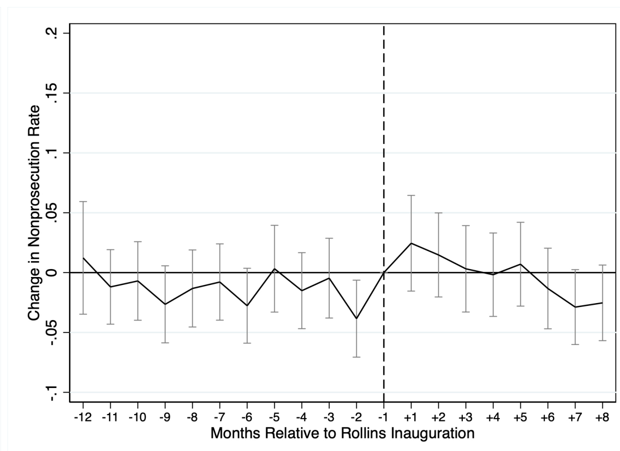
(b) MTE

Figure 4: Marginal Treatment Effects

Notes. In panel (a) the dashed lines represent the upper and lower bounds on the common support of the propensity score (based on 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. The MTE estimation is based on a local IV using a cubic polynomial specification in the sample with common support. The x-axis in panel (b) is the predicted probability of nonprosecution estimated for the assigned ADA 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 a new criminal complaint within two years. The upper right corner of panel (b) shows the estimated LATE, average treatment effect (ATE), average treatment on the untreated (ATUT), and average treatment on the treated (ATT), estimated by rescaling the weights on the MTEs for those parameters to integrate over the common support shown in panel (a) (Carneiro et al., 2011). All estimations were done via `mtefe` in Stata (Andresen, 2019). Appendix Figure B.3 shows results using different specifications.



(a) Nonviolent Misdemeanors



(b) Nonviolent Felonies

Figure 5: Event Study Plots of Nonprosecution Rates, 1/1/2018 - 9/1/2019

Notes. In (a) the sample consists of nonviolent misdemeanor complaints initiated between January 1, 2018 and September 1, 2019; in (b) the sample consists of nonviolent felony complaints initiated over the same period. District Attorney Rachael Rollins was inaugurated on January 2, 2019. Figures report point estimates and 90% confidence intervals for monthly changes in nonprosecution rates relative to December 2018 (the omitted month). Models include court and day-of-week fixed effects and all case and defendant covariates. Robust standard errors clustered at the defendant level.

Table 1: Summary Statistics

	(1) All	(2) Prosecuted	(3) Not Prosecuted
Baseline:			
Not Prosecuted	0.204	0.000	1.000
Number Counts	1.717	1.752	1.581
Number Misdemeanor Counts	1.320	1.365	1.141
Number of Serious Misdemeanor Counts	0.575	0.648	0.290
Misd Conviction within Past Year	0.086	0.100	0.031
Felony Conviction within Past Year	0.045	0.053	0.014
Citizen	0.764	0.743	0.848
Disorderly/Theft	0.284	0.307	0.196
Motor Vehicle	0.394	0.333	0.633
Drug	0.152	0.185	0.027
Other Crime	0.170	0.176	0.144
Victimless Crime	0.817	0.789	0.925
Male	0.799	0.814	0.739
Age $\leq$ 23	0.231	0.233	0.222
Age 24-30	0.246	0.245	0.250
Age 31-40	0.218	0.220	0.210
Age $\geq$ 41	0.306	0.302	0.318
Prob Hispanic	0.332	0.338	0.308
Prob Black	0.345	0.355	0.306
Prob White	0.256	0.242	0.312
Case Outcomes:			
ADA Requested Bail	0.064	0.080	0.000
Bail Set at Arraignment	0.052	0.066	0.000
Days to Disposition	146.930	184.477	0.000
Number of Case Events	3.436	4.058	1.002
DCJIS Record of Case	0.689	0.771	0.366
Any Conviction	0.210	0.263	0.000
Post-Case Outcomes:			
Criminal Complaint Within 2 Years	0.341	0.372	0.217
Number of Complains Within 2 Years	1.467	1.635	0.810
Observations	67060	53411	13649

Notes. This sample includes cases with an arraignment hearing between January 1, 2004 – September 1, 2018, that have no felony or violent/gun misdemeanor charges, that are arraigned in one of Suffolk County's 9 district/municipal courts, that have an identified Assistant District Attorney (ADA) at arraignment, that are processed by an ADA who arraigned at least 30 nonviolent misdemeanor cases, that are not "singletons" within our set of court-by-time fixed effects, and that are not missing gender or age. Source: SCDAO.

Table 2: First Stage: ADA Leniency and Nonprosecution

	(1)	(2)
ADA Leniency	0.60***	0.54***
	(0.07)	(0.07)
Observations	67060	67060
Court x Time FE	Yes	Yes
Case/Def Covariates	No	Yes
Mean Not Prosecuted	0.204	
First Stage F-Stat	66.07	57.94

Notes. This table reports first stage results via a linear probability model for the outcome of nonprosecution. The regressions are estimated on the sample as described in the notes to Table 1. ADA leniency is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. Column (1) reports results controlling for our full set of court-by-time fixed effects. Column (2) adds defendant and case covariates: number of counts; number of misdemeanor counts; number of serious misdemeanor counts; whether the defendant had a prior misdemeanor conviction within the past year; whether the defendant had a prior felony conviction within the past year; indicators for whether the defendant faces charges for a disorder/theft, motor vehicle, drug, or other offense; indicators for citizenship, male, and age categories, and the predicted probability that a defendant is Hispanic, Black, or white. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. Robust (Kleibergen-Paap) first stage F reported (which is equivalent to the effective F-statistic of Montiel Olea and Pflueger (2013) in this case of a single instrument). $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table 3: Second Stage: Probability and Number of Subsequent Criminal Complaints Within Two Years

	OLS		IV		RF
	(1)	(2)	(3)	(4)	(5)
Panel A: Criminal Complaint Within 2 Years					
Not Prosecuted	-0.14***	-0.10***	-0.36***	-0.29***	
	(0.01)	(0.01)	(0.10)	(0.10)	
			[-0.57, -0.15]	[-0.49, -0.07]	
ADA Leniency					-0.16**
					(0.06)
Mean Dep Var Prosecuted	0.37				
Mean Dep Var Prosecuted Compliers	0.55				
Panel B: Number Criminal Complaints Within 2 Years					
Not Prosecuted	-0.73***	-0.51***	-2.27***	-1.71**	
	(0.04)	(0.03)	(0.66)	(0.67)	
			[-3.67, -1.00]	[-3.15, -0.42]	
ADA Leniency					-0.93**
					(0.36)
Mean Dep Var Prosecuted	1.64				
Mean Dep Var Prosecuted Compliers	2.84				
Observations	67060	67060	67060	67060	67060
Court x Time FE	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	No	Yes	No	Yes	Yes

Notes. This table reports OLS and two-stage least squares estimates of the impact of nonprosecution on the probability and number of subsequent criminal complaints within two years, as well as reduced form impacts of leniency on subsequent criminal complaints within two years. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variables are identified in the panel headings. Each panel reports the mean of the dependent variable for all prosecuted defendants, and for prosecuted defendants within the set of compliers. See Appendix A.2 for details on the calculation of mean outcomes among prosecuted compliers. Two-stage least squares models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-year-month and court-by-day-of-week fixed effects; case and defendant covariates are as identified in the notes to Table 2. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses in columns (1)-(4). For the IV estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table 4: Probability of Subsequent Criminal Complaint Within Two Years by Defendant Criminal History

	Prev Complaint		Prev DCJIS		
	(1)	(2)	(3)	(4)	(5)
	No	Yes	No	Yes, No Conv	Yes, Has Conv
Not Prosecuted	-0.18*	-0.04	-0.26***	0.55	0.15
	(0.10)	(0.23)	(0.10)	(0.58)	(0.47)
	[-0.39, 0.04]	[-0.49, 0.51]	[-0.46, -0.05]	[-0.55, 2.95]	[-0.74, 1.41]
Observations	33367	33562	38472	12172	16168
Mean Dep Var Prosecuted	0.20	0.52	0.22	0.46	0.61
Mean Dep Var Prosecuted Compliers	0.26	0.61	0.32	0.37	0.77

Notes. This table reports two-stage least squares estimates of the impact of nonprosecution on on the probability of a subsequent criminal complaint within two years, for first-time and repeat defendants (defined in turn as having any prior complaint in Suffolk County (in column 2); having a prior complaint in Suffolk County that resulted in a DCJIS record but no conviction (in column 4); and having a prior complaint in Suffolk County that resulted in a conviction (in column 5). We report the means of the dependent variable for prosecuted defendants by subsample. See Appendix A.2 for details on the calculation of mean outcomes among prosecuted compliers. The models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-year-month and court-by-day-of-week fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. Confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table 5: Probability of Subsequent Criminal Complaint Within Two Years
By Subsequent Complaint Features

	(1) Non-Discretionary Arrest	(2) Discretionary Arrest	(3) Victim Crime	(4) Victimless Crime	(5) Same Police Agency	(6) Diff Police Agency
Panel A: First Time Defendants						
Not Prosecuted	-0.14**	-0.03	-0.13*	-0.04	-0.10	-0.20*
	(0.07)	(0.09)	(0.07)	(0.08)	(0.11)	(0.11)
	[-0.28, -0.01]	[-0.19, 0.20]	[-0.27, 0.02]	[-0.19, 0.15]	[-0.30, 0.17]	[-0.43, 0.05]
Observations	32981	32981	32981	32981	28371	28371
Mean Dep Var Prosecuted Compliers	0.15	0.11	0.13	0.12	0.12	0.29
Panel B: Non-First Time Defendants						
Not Prosecuted	-0.05	0.00	-0.26	0.21	0.06	-0.12
	(0.22)	(0.25)	(0.21)	(0.24)	(0.22)	(0.29)
	[-0.48, 0.46]	[-0.54, 0.55]	[-0.73, 0.15]	[-0.23, 0.80]	[-0.35, 0.62]	[-0.78, 0.50]
Observations	33498	33498	33498	33498	28931	28931
Mean Dep Var Prosecuted Compliers	0.33	0.29	0.44	0.17	0.32	0.67
Court x Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports two-stage least squares estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within 2 years for different categories of subsequent complaints, as indicated by the column headers, for first-time and repeat defendants. First-time and repeat defendants are defined by reference to any prior criminal complaint in Suffolk County (akin to columns (1)-(2) in Table 4). Non-discretionary and discretionary arrests are defined using Abdul-Razzak and Hallberg (2021), as described in the text. "Victim" offenses include property offenses (e.g., larceny, shoplifting, burglary), threats, property damage, and leaving the scene of property damage or personal injury. "Same police agency" identifies whether the subsequent complaint was brought by the same policy agency as the current complaint. Each panel reports the means of the dependent variable for prosecuted compliers. See Appendix A.2 for details on the calculation of mean outcomes among prosecuted compliers. The models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-year-month and court-by-day-of-week fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. For the IV estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table 6: ADA Missingness: Probability of Subsequent Complaint Within Two Years

	(1)	(2)	(3)	
	Main Sample	Missing ADA Sample	Full Relevant Sample	
Panel A: OLS				
Not Prosecuted	-0.097***	-0.098***	-0.097***	
	(0.005)	(0.003)	(0.002)	
Observations	67060	136974	204034	
Mean Dep Var Pros	0.356	0.371	0.366	
	(1)	(2)	(3)	
	Main	Courts Miss < 50%	Years Miss < 60%	
Panel B: 2SLS in Subsamples				
Not Prosecuted	-0.288***	-0.473**	-0.298	
	(0.105)	(0.199)	(0.230)	
	[-0.493, -0.068]	[-1.136, -0.064]	[-0.799, 0.176]	
Observations	67060	23584	26608	
Mean Dep Var Pros Compliers	0.550	0.622	0.749	
Randomization p	0.169	0.813	0.539	
First-Stage Coef	0.540	0.493	0.424	
First-Stage F	57.94	8.783	27.50	
Prop. Relevant Missing ADA (Pros)	0.640			
Prop. Relevant Missing ADA (Non-Pros)	0.751			
	(1)	(2)	(3)	(4)
	Imputation 1	Imputation 2	Imputation 3	Imputation 4
Panel C: 2SLS in Imputation Samples				
Not Prosecuted	-0.270***	-0.285**	-0.278***	-0.310***
	(0.087)	(0.124)	(0.095)	(0.100)
	[-0.442, -0.092]	[-0.524, 0.030]	[-0.462, -0.045]	[-0.522, -0.077]
Observations	85015	111709	132905	147080
Mean Dep Var Pros Compliers	0.503	0.663	0.613	0.671
Randomization p	0.0520	0.250	0.206	0.409
First-Stage Coef	0.537	0.320	0.351	0.312
First-Stage F	80.54	18.19	20.36	17.38
Prop. Relevant Missing ADA (Pros)	0.564	0.403	0.309	0.241
Prop. Relevant Missing ADA (Non-Pros)	0.625	0.448	0.311	0.224
Court x Time FE	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes

Notes. This table addresses missing data on the identity of the arraigning ADA. Panel A reports OLS estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within two years, for alternative samples of cases: column (1) is our main sample; column (2) has the same restrictions as our main sample but is additionally restricted to cases missing ADA information; column (3) combines columns (1) and (2). In Panel B, column (2) replicates our main 2SLS analysis within courts where arraigning ADA information is missing less than 50% of the time (South Boston, East Boston, and West Roxbury). Column (3) replicates this analysis within years where arraigning ADA information is missing less than 60% of the time (2004, 2006-2008). Finally, columns (1) - (4) of Panel C report 2SLS estimates for progressively expanded samples of cases for which arraigning ADA assignment has been imputed following the strategies described in Section C.2. All models instrument for nonprosecution using our main ADA leniency measure, estimated using only cases assigned to an observed arraigning ADA following the procedure described in the text. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. Confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table 7: Effect of Rollins Inauguration on Probability of Subsequent Criminal Complaint Within One Year

	(1) OLS	(2) IV	(3) IVDD NV Fel Control	(4) RF	(5) RF NV Fel Control
Panel A: Dependent Var: Criminal Complaint within 1 Year					
Not Prosecuted	-0.08***	-0.59***	-0.52*		
	(0.01)	(0.14)	(0.29)		
Post Rollins			0.00	-0.03***	-0.00
			(0.02)	(0.01)	(0.01)
Post Rollins x NV Misd					-0.03*
					(0.01)
Observations	19502	19502	25559	19502	25559
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes
Court & Time FE	Yes	Yes	Yes	Yes	Yes
Mean Prosecuted	0.301				
Mean Prosecuted Compliers		0.361	0.361		
Panel B: First-Stage, Dependent Var: Not Prosecuted					
Post Rollins		0.06***	0.01		
		(0.01)	(0.01)		
Post Rollins x NV Misd			0.05***		
			(0.01)		
Case/Def Covariates		Yes	Yes		
Court & Time FE		Yes	Yes		
First Stage F-Stat		58.44	28.57		
Mean NV Misd (Pre-Rollins)		0.355	0.355		
Mean NV Felony (Pre-Rollins)			0.0534		

Notes. This table reports OLS, two-stage least squares, and reduced form OLS estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within one year. The regressions are estimated on the sample of cases involving nonviolent misdemeanor complaints initiated between January 1, 2018 and September 1, 2019. Two-stage least squares and reduced form models instrument for nonprosecution using an indicator for the post-Rollins period. Models (3) and (5) include as a control group the sample of cases involving nonviolent felony complaints initiated between January 1, 2018 and September 1, 2019. The dependent variable in Panel A is an indicator for whether a defendant receives a new criminal complaint within one year post-arraignment; the dependent variable in Panel B is an indicator for whether a defendant's criminal complaint is not prosecuted. Each panel reports the mean of the dependent variable for all prosecuted defendants, and for prosecuted defendants within the set of compliers. See Appendix A.2 for details on the calculation of mean outcomes among prosecuted compliers. All specifications include all case and defendant covariates and month and day-of-week fixed effects. Robust standard errors clustered on defendant are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Appendix for:

Misdemeanor Prosecution

Amanda Agan, Jennifer Doleac, and Anna Harvey

A Technical Appendix

A.1 Comparisons to Alternative Instrument Estimation Strategies

Our main instrument is a residualized leave-out mean leniency measure that is estimated from the other nonviolent misdemeanor cases that an ADA has arraigned. For our main analyses we proceed by implementing this instrument in an 'as-if just-identified' manner: we report robust F-statistics that do not adjust for the fact that the instrument is estimated (although we also do not use these F-statistics directly in a threshold test), and we conduct inference in the second stage using confidence intervals based on inversion of the Anderson-Rubin test, which have correct size and optimal power even when instruments are weak in just-identified models (Anderson and Rubin, 1949; Andrews et al., 2019). Performing the estimation in this way is standard (Doyle Jr, 2007; French and Song, 2014; Dahl et al., 2014; Dobbie et al., 2018; Bhuller et al., 2020) and has some attractive properties. We have 315 ADAs in our sample, each of whom could serve as a potential instrument. With this many instruments, estimation using the ADA dummies can suffer from bias from many (potentially weak) instruments (Bekker, 1994; Bound et al., 1995; Hausman et al., 2012). We also have many fixed effects (court-by-year-month and court-by-day of week) in the covariate set, necessary to identify the set of cases for which ADAs are as-if randomly assigned, which can also cause bias in jackknife instrumental variable estimators (JIVE) (Kolesár, 2013). It is also clear in the just-identified case how to handle inference in the second stage that is robust to (potential) weak instrument issues (Andrews et al., 2019). The continuous instrument also allows for the estimation of marginal treatment effects (Heckman and Vytlacil, 2005).

While this estimation strategy is convenient for the reasons mentioned above, it does not take into account that the instrument itself is constructed. In this subsection we explore the robustness of our main estimates to alternative estimation strategies in this setting. One alternative approach is to estimate our 2SLS model using the full set of 315 ADA dummy variables as instruments in the first stage. These results are shown in Table A.1, column (2) (column (1) repeats our main 2SLS results using the residualized leave-out mean leniency as an instrument). The estimated coefficient is -0.19, smaller in absolute value and closer to OLS than our main leave-out mean 2SLS estimate, which is unsurprising given that the bias from weak instruments moves estimates closer to the OLS estimate. Several strategies are suggested when IV estimates suffer from bias from many (weak) instruments. In column (3) we estimate a limited information maximum likelihood (LIML) model with all the dummies as instruments (Bekker, 1994; Chao and Swanson, 2005; Angrist and Frandsen, 2020). The coefficient in this model is -0.27, closer to our main leave-out mean 2SLS estimate than the estimate in column (2) with all the ADA dummy variables. LIML however is not consistent with heteroskedastic errors or with heterogeneous treatment effects (Hausman et al., 2012; Kolesár, 2013), which is the motivation for the UJIVE estimator we also use. In column (4) we estimate the unbiased JIVE (UJIVE) estimator of Kolesár (2013).[23] JIVE estimators are generally suggested when the number of instruments is large (Angrist et al., 1999), although they can be biased with many covariates (Kolesár, 2013). The UJIVE estimator is consistent (for a convex combination of LATE estimates) with a large number of covariates. Here the coefficient is -0.20, again closer to our main leave-out mean 2SLS estimate than the estimate in column (2) with all the ADA dummy variables. Goldsmith-Pinkham et al. (2022) point out that linear regressions with multiple treatments and flexible controls may not recover convex averages; our linear first stage with multiple ADA dummies may be contaminated with non-convex averages of other ADAs (treatments). To address this concern they suggest interacting the individual ADA instrument with our court x time fixed effects. Similarly, Blandhol et al. (2022) note that when 2SLS relies on covariates, a "saturated and weighted" specification may be necessary to recover properly weighted averages of covariate-specific LATEs. Column (5) reports these estimates. The coefficient is smaller (-0.15) but still

[23]We thank Paul Goldsmith-Pinkham, Peter Hull, Michal Kolesár, and Mauricio Bravo for sharing the Stata command `manyiv` to calculate the UJIVE results.

negative, statistically significant, and large compared to the complier means (27% decline in recidivism). Another way to handle the potential bias from many (weak) instruments is to reduce the number of instruments by using lasso to pick the most informative ADA dummies in a 2SLS regression. We do this in column (6) using a post-lasso first stage via the procedures of Belloni et al. (2014).[24]

In each case the coefficients we estimate with the alternative strategies are negative, large, and statistically significant, implying that nonprosecution decreases criminal complaints within two years post-arraignment between 33-70% relative to prosecuted compliers. For the most part, given the estimated standard errors, we cannot reject the null that these coefficients are the same as our main 2SLS estimate.[25]

[24] In practice this was 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. The procedure retains three out of 315 instruments (similarly in the Angrist and Frandsen (2020) implementation of the plug-in penalty, lasso retains two instruments out of 180 in a re-estimation of the Angrist and Keueger (1991) QOB study). We also implemented a version of `ivlasso` with a cross-validated penalty; see Angrist and Frandsen (2020) for details on implementation. The algorithm with the CV penalty chooses more instruments, namely 173 out of the 315 in our case. The estimated post-lasso coefficient is smaller in absolute value (-0.22, se=0.049). The simulation results of Angrist et al. (1999) and Belloni et al. (2012) imply that the plug-in penalty will have less bias although will also be less precise than the CV penalty estimates.

[25] The standard errors reported here for 2SLS or LIML using all the dummy variables have not been adjusted to take into account the potential for weak instruments; other inference strategies may imply larger confidence intervals.

Table A.1: Different IV Estimation Strategies

	(1)	(2)	(3)	(4)	(5)	(6)
	Main	All Dummies	LIML	UJIVE	UJIVE Interacted	lasso
Not Prosecuted	-0.29***	-0.16***	-0.22**	-0.20**	-0.15***	-0.32***
	(0.10)	(0.05)	(0.09)	(0.08)	(0.04)	(0.11)
Observations	67060	67060	67060	67060	60554	67060
Court x Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Mean Not Prosecuted	0.372					
Mean Not Prosecuted Compliers	0.550					

Notes. This table reports two-stage least squares estimates using various estimation strategies for the instrument, as indicated in the column headers. All specifications control for court-by-time fixed effects and case/defendant covariates. The OLS estimate for this specification can be found in Table 3 column (2), and is -0.10 (se=0.01). Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses (except for UJIVE where they are only clustered at the individual defendant level, as the code only allows for one cluster; clustering on the ADA level instead does not change results). Column (1) repeats our main 2SLS estimates using the residualized leave-out mean leniency measure with covariates (see Table 3, column (4)). In column (2) we use all 315 ADA dummy variables directly as instruments in the first stage. Column (3) uses limited information maximum likelihood estimation with all of the dummies as instruments. Column (4) uses the UJIVE estimator of Kolesár (2013). Column (5) repeats the UJIVE estimation but the instruments are now ADAs interacted with our court x time fixed effects as suggested by Goldsmith-Pinkham et al. (2022). Column (6) uses post-lasso from Belloni et al. (2014) to choose the most informative ADA dummy variables; the algorithm chooses three of the ADA dummies as instruments. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

A.2 Understanding Compliers

To calculate the shares of compliers, never-takers, and always-takers, we use the insights of Abadie (2003) and Dahl et al. (2014), and applied by Dobbie et al. (2018) and Bhuller et al. (2020).

Always-takers are defendants who would always be not prosecuted regardless of the ADA assigned to their case. Given our monotonicity and conditional independence assumptions, the fraction of always-takers can be calculated by the share of defendants not prosecuted by the most strict ADA(s):

$$\pi_a = Pr(\text{Not Prosecuted}_i = 1 | Z_i = \underline{z}) = Pr(\text{Not Prosecuted}_i(\overline{z}) = \text{Not Prosecuted}_i(\underline{z}) = 1) \tag{4}$$

where $\overline{z}$ represents a maximum value of the ADA instrument (the most lenient ADA) and $\underline{z}$ represents a minimum value of the instrument (the most strict ADA).

Similarly, never-takers are defendants who would never be not prosecuted (always be prosecuted). We can estimate their fraction by the share of defendants who are prosecuted by the most lenient ADA(s):

$$\pi_n = Pr(\text{Not Prosecuted}_i = 0 | Z_i = \overline{z}) = Pr(\text{Not Prosecuted}_i(\overline{z}) = \text{Not Prosecuted}_i(\underline{z}) = 0) \tag{5}$$

Finally, compliers are defendants whose prosecution decisions would have been different had their case been assigned to the most lenient instead of the most strict ADA:

$$\begin{aligned}\pi_c = Pr(\text{Not Prosecuted}_i = 1 | Z_i = \overline{z}) - Pr(\text{Not Prosecuted}_i = 1 | Z_i = \underline{z}) = \\ Pr(\text{Not Prosecuted}_i(\overline{z}) > \text{Not Prosecuted}_i(\underline{z}))\end{aligned} \tag{6}$$

We can calculate this as $1 - \pi_a - \pi_n$. Under a linear specification of the first stage Equation 3, we can recover π_c as $\alpha_1(\overline{z} - \underline{z})$, π_a as $\alpha_0 + \hat{\alpha_1}\underline{z}$, and π_n as $1 - \alpha_0 - \hat{\alpha_1}\overline{z}$, where α_0 and α_1 are the estimated first stage coefficients. We also estimate these under more flexible local linear estimations of our first stage.

We define most and least lenient ADAs by their percentiles in the residualized leniency distribution, defining the least lenient ADAs as those at the ρ percentile and the most lenient as those at the $(100 - \rho)$ percentile, where ρ varies between 1, 1.5, and 2. In the first three columns of Appendix Table B.10, we use a linear specification of the first stage, given by Equation 3. Under this linear specification, we find that ten percent of our sample are compliers, 73 percent are never takers, and 18 percent are always takers. The latter three columns use a local linear version of our first stage of nonprosecution on the residualized measure of ADA leniency, controlling for court-by-time fixed effects. Under this more flexible analog to our first stage equation, we find that nine percent of our sample are compliers, 72 percent are never-takers, and 18 percent are always-takers.

We then use a similar insight to describe observable characteristics of compliers by calculating the fraction of compliers in different subsamples (Abadie 2003). With the shares calculated above, we can then calculate average characteristics for complier defendants who were prosecuted: $E(Y_i(0)|\text{Not Prosecuted}_i(\overline{z}) > \text{Not Prosecuted}_i(\underline{z}))$. Among the prosecuted, average outcomes for defendants who were assigned to lenient ADAs are average outcomes for the never-takers:

$$E(Y_i|\text{Not Prosecuted}_i = 0, z_i = \overline{z}) = E(Y_i(0)|\text{Not Prosecuted}_i(\overline{z}) = \text{Not Prosecuted}_i(\underline{z}) = 0) \tag{7}$$

Among the prosecuted, outcomes for defendants who were assigned to strict ADAs are a weighted average of outcomes for compliers and never-takers, where the weights are their shares in the population:

$$E(Y_i|NotProsecuted_i = 0, z_i = \underline{z}) = \frac{\pi_c}{\pi_c + \pi_n} E(Y_i(0)|\text{Not Prosecuted}_i(\overline{z}) > \text{Not Prosecuted}_i(\underline{z})) + \frac{\pi_n}{\pi_c + \pi_n} E(Y_i(0)|\text{Not Prosecuted}_i(\overline{z}) = \text{Not Prosecuted}_i(\underline{z}) = 0) \tag{8}$$

Plugging Equation 7 into Equation 8, we can calculate average outcomes for compliers among the prosecuted for any outcome Y_i as:

$$E(Y_i(0)|\text{Not Prosecuted}_i(\overline{z}) > \text{Not Prosecuted}_i(\underline{z})) = \frac{\pi_c + \pi_n}{\pi_c} E(Y_i|\text{Not Prosecuted}_i = 0, z_i = \underline{z}) - \frac{\pi_n}{\pi_c} E(Y_i|\text{Not Prosecuted}_i = 0, z_i = \overline{z}) \tag{9}$$

Appendix Table B.11 shows these results for various observable characteristics.[26] In column (1) we show the proportion of our sample represented by this subset of observable characteristics; in column (2) we show the estimated proportion of this subsample composed of compliers; in column (3) we show the ratio of how often the trait occurs in the estimated complier group, relative to the full sample. Compliers look similar to the full sample on many dimensions, but differ on others. In particular, compliers are less likely to have been charged with a drug offense, to have been charged with a serious misdemeanor (punishable by more than 100 days in jail), to have misdemeanor or felony convictions within the prior year, and to be noncitizens, and more likely to be younger (less than 24 years old) and female.

[26] For these calculations we use the linear specification of the first stage and a $\rho = 1\%$ cutoff to define most and least lenient ADAs.

A.3 Empirical Bayes Shrinkage

Our prosecutor leniency measure is based on a finite number of observations of arraignment decisions made by individual ADAs. We restrict our sample to observations for ADAs who have made decisions on at least 30 arraignments to try to remove some of the finite sample bias. But it is still true that each measure of ADA leniency has error, and the fewer observations we have on an ADA, the more error the measure may contain. We do not want that finite-sample error to be driving the results. This is similar to issues in measuring teacher value-added when one only has a finite number of observations (see e.g. Chetty et al. 2014). The common resolution in that literature is to use empirical Bayes shrinkage procedures to shrink the fixed effects towards a prior mean–with noisier estimates being "shrunk" more, achieved by multiplying our estimated leniency by a reliability/shrinkage factor based on how noisy the individual ADA's leniency estimate is compared to the distribution of all leniency measures, or the ratio of signal variance to total variance (Morris 1983; Kane and Staiger 2008; Chetty et al. 2014; Arnold et al. 2020; Rivera 2021).

To get the shrinkage factor using empirical Bayes shrinkage we follow the literature cited above. Assume the (true) leniency measures θ_j are drawn from a normal distribution $N(0, \sigma_\theta^2)$. For each prosecutor we see a noisy estimate of true leniency, $\hat{\theta}_j = \theta_j + \epsilon_j$, with $\epsilon_j \sim N(0, \epsilon_j)$. Given a prior mean of 0 (given our residualized leave-out measure), the posterior mean is $E(\theta_j|\hat{\theta}_j) = \hat{\theta}_j \frac{\sigma^2}{\sigma^2 + \frac{\sigma_i^2}{n}}$, where $\hat{\theta}_j$ is our estimated residualized leniency measure (Z_{cta} in the notation of section 3); σ^2 is the variance across estimates of $\hat{\theta}_i$; and $\frac{\sigma_i^2}{n}$ is the square of the within prosecutor standard error of residualized leniency estimates ($NotProsecuted^*_{ict}$ in Section 3).

Appendix Figure B.1 shows how the distribution of shrunk leniency estimates compares to our main leniency measure. Appendix Table B.9 column (3) reports 2SLS estimates using the shrunk leniency estimates rather than our main leniency estimates.

A.4 MTE Estimation

Re-orienting our framework to the potential outcome framework, let $Y_i(1)$ be the defendant's outcome Y if not prosecuted ($D_i = 1$) and $Y_i(0)$ the defendant's outcome if prosecuted ($D_i = 0$). An ADA makes a decision to prosecute or not prosecute a defendant based on characteristics both observable to the econometrician, X_i, and unobservable to the econometrician, ν_i. Define the latent propensity to be not prosecuted as: $D_i^* = \mu_D(Z_i, X_i) - \nu_i$. $D_i = 1$, or the defendant is not prosecuted, if $D_i^* = \mu_D(Z_i, X_i) - \nu_i \geq 0 \implies \mu_D(Z_i, X_i) \geq \nu_i \implies F_\nu(\mu_D(Z_i, X_i)) \geq F_\nu(\nu_i)$ and prosecuted otherwise, where F_ν is the (unknown) cumulative distribution function of ν. $F_\nu(\mu_D(Z_i, X_i)) = P(Z_i, X_i)$ is the propensity score: the probability of nonprosecution conditional on observables, X_i, and ADA leniency, Z_i. Call $F_\nu(\nu_i) = U_D$ quantiles of the distribution of the unobserved resistance (since ν enters the selection equation with a negative sign) to be not prosecuted. The marginal treatment effect is then defined as the treatment effect at a particular value of the unobservable propensity to be not prosecuted: $E(Y_i(1) - Y_i(0)|X_i = x, U_{Di} = u_D)$, that is, the treatment effect for individuals on the margin of being not prosecuted when $P(Z_i, X_i) = u_D$. It can be estimated as the derivative of the average outcome conditional on X and $P(Z_i, X_i) = u_D$ with respect to the propensity score.

B Additional Figures and Tables

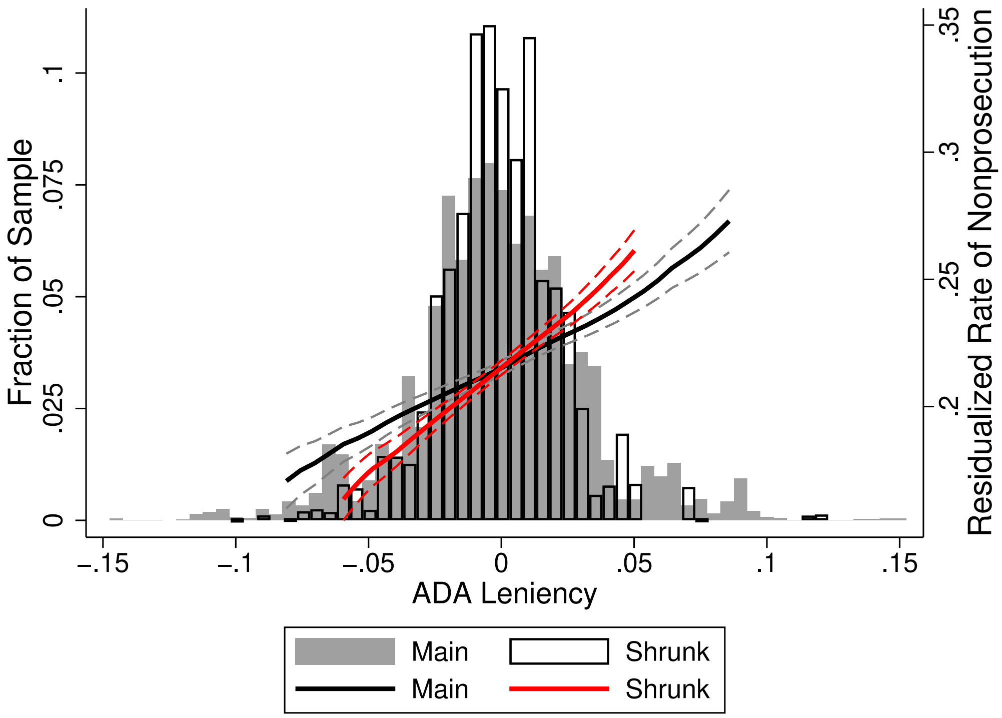


Figure B.1: ADA Leniency and Nonprosecution:
Main and Empirical Bayes Shrinkage

Notes. The "Main" estimates repeat Figure 1, showing the distribution of our main leave-out mean measure of ADA "leniency," residualized by court-by-year-month and court-by-day-of-week; the solid black line is a local linear regression of nonprosecution on ADA leniency, along with the 95% confidence interval, estimated from the 1st to 99th percentiles of ADA leniency—a local linear version of our first stage. The "shrunk" estimates apply an empirical Bayes shrinkage factor to our main leniency estimates to account for potential finite sample bias—the leniency measures are shrunk towards the prior mean based on how noisy the individual ADA's leniency estimate is compared to the distribution of all leniency measures (see Appendix Section A.3 for details). The standard deviation of our main ADA leniency estimate is 0.04 and of the shrunken estimates is 0.026.

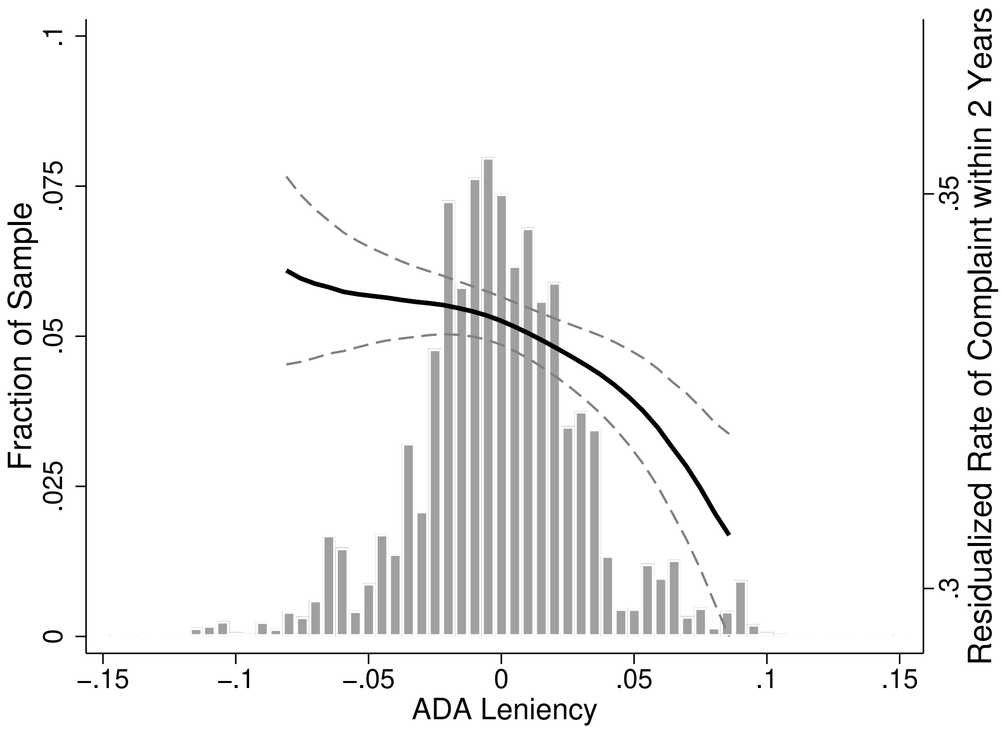


Figure B.2: ADA Leniency and Probability of
Subsequent Criminal Complaint within 2 Years

Notes. This figure shows the distribution of our leave-out mean measure of ADA "leniency," residualized by court-by-year-month and court-by-day-of-week, as in Figure 1. The solid line is a local linear regression of whether the defendant has a subsequent criminal complaint within 2 years on ADA leniency, along with the 95% confidence interval, estimated from the 1st to 99th percentiles of ADA leniency—a local linear version of our reduced form. A defendant assigned to a more lenient ADA (computed using all cases except the current case and other cases with the same defendant) has a lower likelihood of a subsequent criminal complaint.

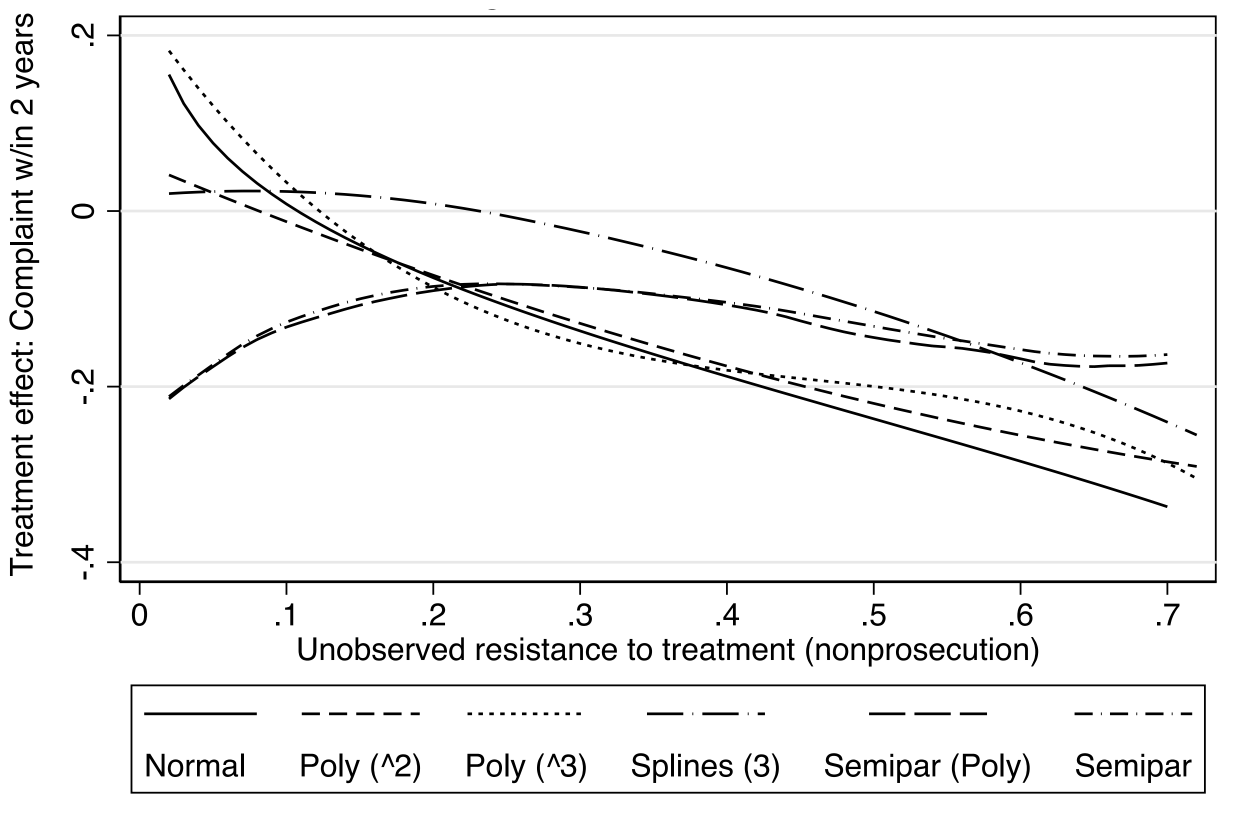


Figure B.3: Marginal Treatment Effects: Different Modelling Assumptions

Notes. This figure shows the estimated MTEs under various parametric and semi-parametric modeling assumptions. Propensity scores are predicted via a logit regression with all case- and defendant-level covariates included, including court-by-time fixed effects. The MTE estimation is based on a local IV, using one of the modeling assumptions listed in the legend in the sample with common support (see Figure 4a). The cubic polynomial model is the model used in Figure 4b and is replicated here for comparison. The modeling assumptions are (in the order of the legend): a joint normal model, a quadratic polynomial, a cubic polynomial, a quadratic polynomial with 3 splines (at 0.25, 0.5, and 0.75), a semi-parametric polynomial model, and a semi-parametric local IV model; these are described in more detail in andresen2018exploring. All estimations were done via `mtefe` in Stata (mtefe).

(a) Common Support

(b) MTE

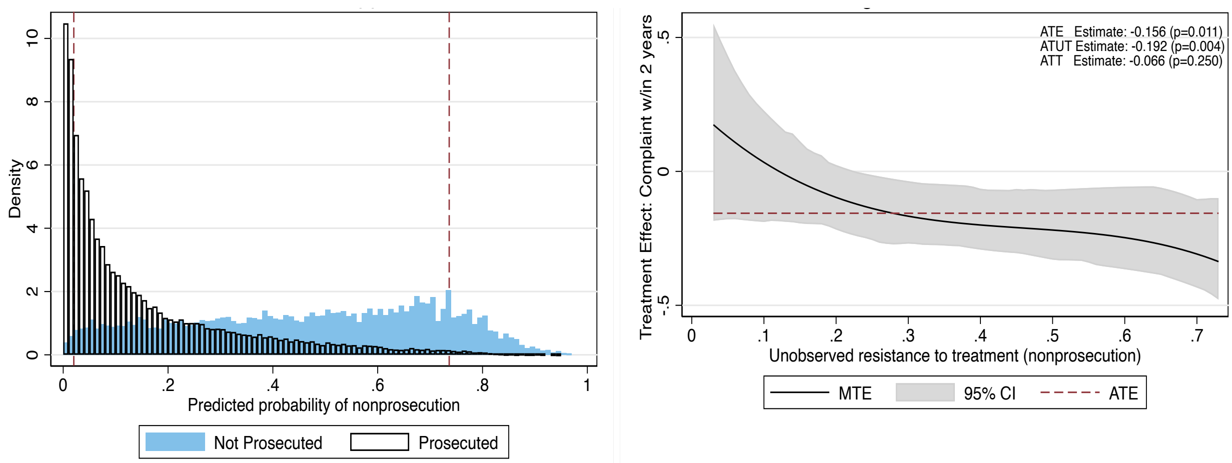


Figure B.4: Marginal Treatment Effects:
Within "Strict Monotonicity" Courts

Notes. This figure repeats the analysis in Figure 4 with the sample restricted to the 6 courts in which we could not reject the null of strict monotonicity (see Table B.3). In (a) the dashed lines represent the upper and lower bounds on the common support of the propensity score (based on 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. The MTE estimation is based on a local IV using a cubic polynomial specification in the sample with common support. The x-axis in Figure (b) is the predicted probability of nonprosecution estimated from the assigned ADA 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 a new criminal complaint within two years. The upper right corner on Panel (b) shows the estimated average treatment effect (ATE), average treatment on the untreated (ATUT), and average treatment on the treated (ATT). These were estimated by rescaling the weights on the MTEs for those parameters to integrate over the common support shown in (a) (carneiro2011estimating). All estimations were done via `mtefe` in Stata (mtefe).

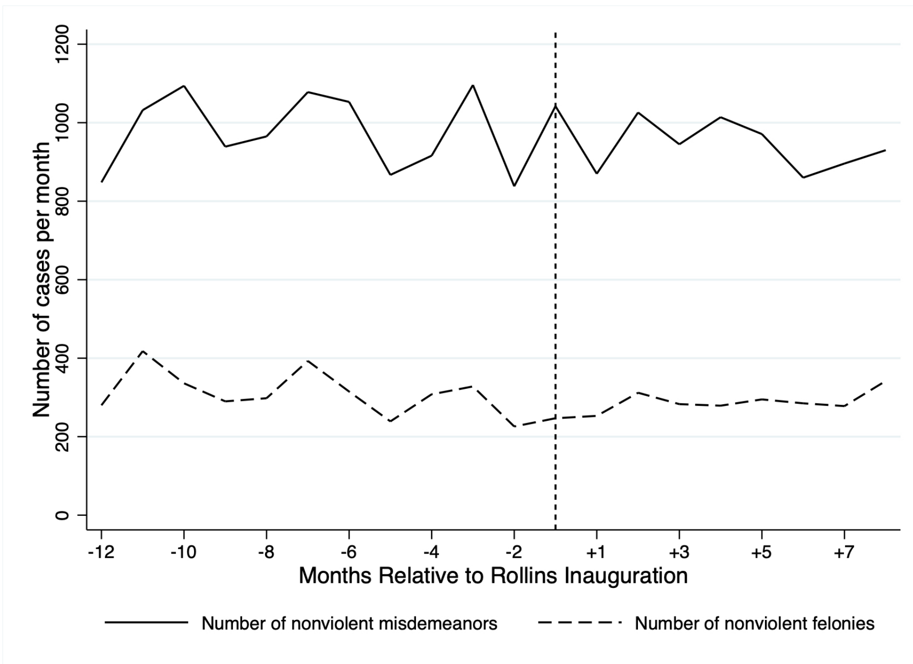
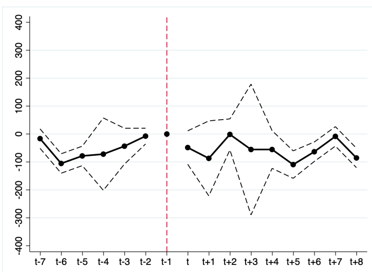
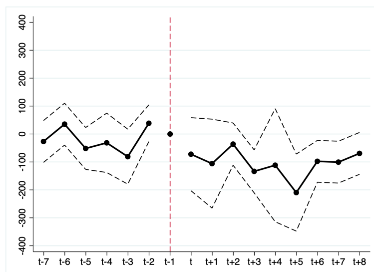


Figure B.5: Numbers of Nonviolent Misdemeanor and Felony Cases

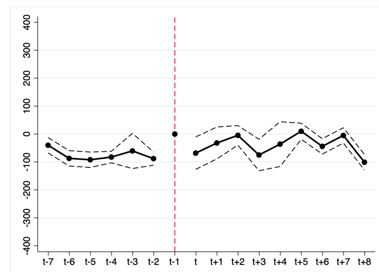
Notes. This figure reports the monthly numbers of cases in Suffolk County involving nonviolent misdemeanor and nonviolent felony complaints initiated between January 1, 2018 and September 1, 2019. District Attorney Rachael Rollins was inaugurated on January 2, 2019.



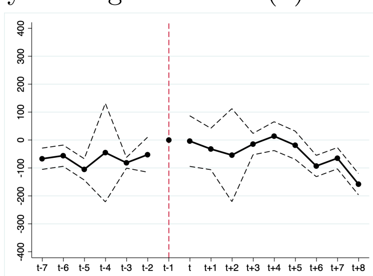
(a) Property Damage



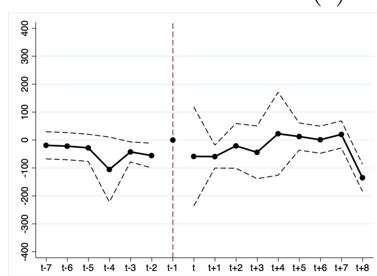
(b) Theft & Fraud



(c) Disorder



(d) Drug

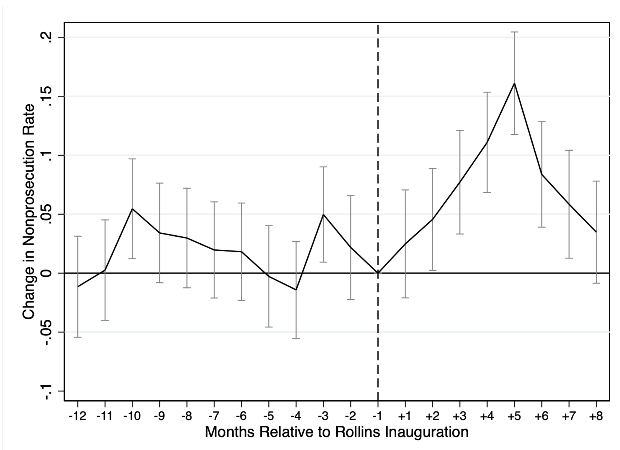


(e) Other

Figure B.6: Event Study Plots of Arrest Rates

Notes. This figure reports point estimates and 95% confidence intervals for monthly changes in the number of arrests made by the Boston Police Department between January 1, 2017 and February 29, 2020. Monthly arrest data are sourced from the FBI's Uniform Crime Reporting program. Each subfigure is a coefficient plot for a particular category of arrests. The y-axes show number of arrests; the x-axes show time (month) relative to December 2018 (the omitted month). District Attorney Rachael Rollins was inaugurated on January 2, 2019. The coefficient for *t-7* includes months June 2018 and earlier; *t+8* includes months September 2019 and later. All regressions include month-of-year fixed effects and estimate robust standard errors.

(a) Nonviolent Misd Rollins on Memo



(b) Nonviolent Misd Not on Memo

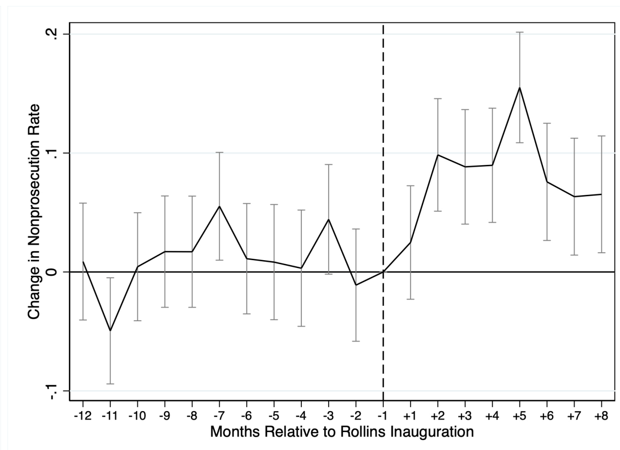
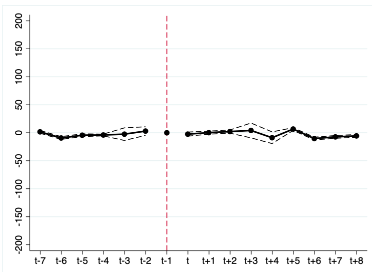
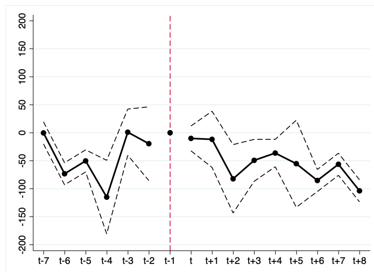


Figure B.7: Additional Event Study Plots of Nonprosecution Rates

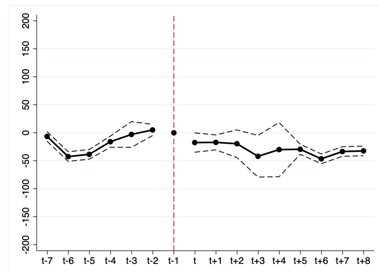
Notes. In (a) the sample consists of nonviolent misdemeanor complaints identified for presumptive nonprosecution in the March 25, 2019 SCDAO policy memo and initiated between January 1, 2018 and September 1, 2019; in (b) the sample consists of nonviolent misdemeanor complaints not identified in this memo initiated over the same period. District Attorney Rachael Rollins was inaugurated on January 2, 2019. Figures report point estimates and 95% confidence intervals for monthly changes in nonprosecution rates relative to December 2018 (the omitted month). Models include court and day-of-week fixed effects and case and defendant covariates: number of counts; number of misdemeanor counts; number of serious misdemeanor counts; whether the defendant had a prior misdemeanor conviction within the past year; whether the defendant had a prior felony conviction within the past year; indicators for whether the defendant faces charges for a disorder/theft, motor vehicle, drug, or other offense; indicators for citizenship, male, and age categories, and predicted probability that a defendant is Hispanic, Black, or white. Robust standard errors clustered at the defendant level.



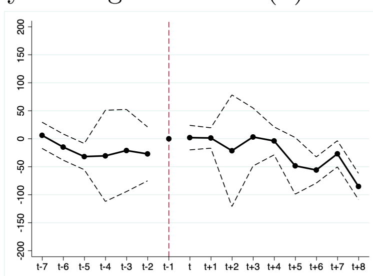
(a) Property Damage



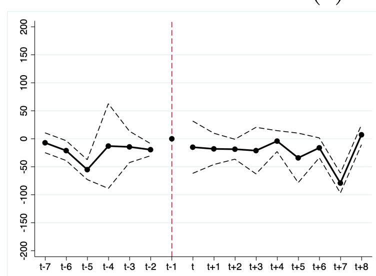
(b) Theft & Fraud



(c) Disorder



(d) Drug



(e) Other

Figure B.8: Event Study Plots of Crime Rates

Notes. This figure reports point estimates and 95% confidence intervals for monthly changes in the number of crime incidents reported by the Boston Police Department between January 1, 2017 and February 29, 2020. Each subfigure is a coefficient plot for a particular category of reported crime. The y-axes show number of reported incidents; the x-axes show time (month) relative to December 2018 (the omitted month). District Attorney Rachael Rollins was inaugurated on January 2, 2019. The coefficient for *t-7* includes months June 2018 and earlier; *t+8* includes months September 2019 and later. All regressions include month-of-year fixed effects and estimate robust standard errors. Data are sourced from https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system. There was a marked increase in the number of "verbal disputes" in the reported crime data beginning in October 2019, a near-doubling of such incidents. Because this seems to be driven by some other factor, we exclude this crime category from the analysis.

Table B.1: Randomization

	(1) Nonprosecution	(2) ADA Leniency (St Dev Units)
Number Counts	-0.017***	-0.002
	(0.003)	(0.009)
Number Misdemeanor Counts	0.016***	-0.001
	(0.004)	(0.013)
Number of Serious Misdemeanor Counts	-0.101***	-0.010
	(0.006)	(0.006)
Misd Conviction within Past Year	-0.061***	-0.014
	(0.005)	(0.010)
Felony Conviction within Past Year	-0.050***	-0.013
	(0.006)	(0.017)
Citizen	0.037***	-0.006
	(0.004)	(0.008)
Disorderly/Theft	-0.021**	-0.016
	(0.008)	(0.016)
Motor Vehicle	0.102***	-0.003
	(0.009)	(0.012)
Drug	-0.093***	-0.020
	(0.008)	(0.013)
Male	-0.060***	-0.030***
	(0.004)	(0.011)
Age 24-30	-0.019***	0.012
	(0.005)	(0.012)
Age 31-40	-0.024***	0.003
	(0.005)	(0.011)
Age $\geq$ 41	-0.011**	0.017
	(0.005)	(0.011)
Prob Hispanic	-0.075***	-0.025
	(0.014)	(0.038)
Prob Black	-0.073***	-0.055
	(0.012)	(0.045)
Prob White	-0.037***	-0.037
	(0.013)	(0.040)
Observations	67060	67060
Joint F-Test p-value	0	0.169

Notes. This table reports regressions testing the random assignment of cases to arraigning ADAs. Column (1) reports estimates from an OLS regression of nonprosecution on the variables listed and court-by-time fixed effects. Column (2) reports estimates from an OLS regression of standardized ADA leniency on the variables listed and court-by-time fixed effects (dependent variable is in standard deviation units). Robust standard errors two-way clustered at the individual and ADA 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 ADA level. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table B.2: Monotonicity

	(1)	(2)	(3)	(4)	(5)	(6)
	1 Count	> 1 Count	Misd 1 Count	Misd > 1 Count	No Misd Conv w/in 1 Year	Misd Conv w/in 1 Year
Not Prosecuted	0.60***	0.57***	0.62***	0.46***	0.48***	0.62***
	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
Observations	36788	30129	48173	18755	31101	63121
	(7)	(8)	(9)	(10)	(11)	(12)
	No Felony Conv w/in 1 Year	Felony Conv w/in 1 Year	Non-Citizen	Citizen	Any Disorder Crime	Any MV Crime
Not Prosecuted	0.61***	0.37*	0.24***	0.69***	0.64***	0.55***
	(0.07)	(0.20)	(0.08)	(0.08)	(0.14)	(0.12)
Observations	65369	1394	15697	51206	22107	30285
	(13)	(14)	(15)	(16)	(17)	(18)
	Any Drug Crime	Any Other Crime	Victimless Crime	Victim Crime	Black Admin Data	Hispanic Admin Data
Not Prosecuted	0.10*	0.42*	0.65***	0.13*	0.54***	0.44***
	(0.06)	(0.21)	(0.08)	(0.08)	(0.08)	(0.17)
Observations	14104	7997	54762	12182	25824	8943
	(19)	(20)	(21)	(22)	(23)	(24)
	White Admin Data	Race Miss Admin Data	Male	Female	Age≤25	Age≥25
Not Prosecuted	0.51***	0.65***	0.55***	0.74***	0.74***	0.52***
	(0.15)	(0.17)	(0.07)	(0.14)	(0.09)	(0.10)
Observations	20447	10293	53550	13383	20596	46350
	(25)	(26)	(27)	(28)		
	Predicted White	Predicted Black	Predicted Hispanic	Predicted Other		
Not Prosecuted	0.65***	0.62***	0.57***	0.63***		
	(0.17)	(0.07)	(0.10)	(0.13)		
Observations	10293	26759	24376	14323		

Notes. This table reports first stage results by subsamples based on case and defendant characteristics, as listed in the column headers. The regressions are estimated on the sample as described in the notes to Table 1. All specifications control for court-by-time fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$

Table B.3: Frandsen, Lefgren, Leslie (2020) Test of Joint Null of Exclusion and Monotonicity, by Court

	Count	FLL p-value
Dorchester	15467	0.107
Roxbury	15055	0.167
Central	10335	0.000
West Roxbury	9514	0.460
East Boston	7757	0.639
South Boston	6385	0.517
Chelsea	1223	0.000
Brighton	1177	0.000
Charlestown	577	0.265

Notes. This table presents results from the test proposed in FrandsenEtAl for the joint null hypothesis that the monotonicity and exclusion restrictions hold. We test this null within courts using day-of-week and year-month fixed effects along with our main covariates. A failure to reject the null implies that we 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 B.4: ADA Leniency and Post-Arraignment Case Outcomes

	Case Outcomes				
	(1) Num Events	(2) Days to Disp	(3) Bail Set	(4) Conviction	(5) DCJIS Match
Panel A: Only Prosecuted Defendants					
ADA Leniency	-0.35	14.45	-0.08*	-0.04	
	(0.43)	(36.51)	(0.04)	(0.07)	
Observations	53373	53373	53373	53373	
Mean Dep Var	4.06	184.48	0.07	0.26	
F-Stat	0.67	0.16	3.34	0.41	
Panel B: Whole Sample					
ADA Leniency	-1.71***	-85.46***	-0.09**	-0.15**	-0.29***
	(0.51)	(32.79)	(0.04)	(0.06)	(0.08)
Observations	67060	67060	67060	67060	67056
Mean Dep Var	3.44	146.93	0.05	0.21	0.69
F-Stat	11.04	6.79	5.76	6.55	13.60
Court x Time FE	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes

Notes. This table reports OLS estimates of the association between ADA leniency and post-arraignment case outcomes, for our whole sample and only among prosecuted defendants. We do not estimate the impact of ADA leniency on DCJIS record acquisition among prosecuted defendants because all prosecuted defendants acquire DCJIS records; variation in this outcome among prosecuted defendants is solely due to measurement error. These models have a similar estimation strategy as the first-stage estimates in Table 2. All specifications control for court-by-year-month and court-by-day-of-week fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses.. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table B.5: Probability of Subsequent Criminal Complaint: Diff Time Horizons

	OLS		IV		RF
	(1)	(2)	(3)	(4)	(5)
Panel A: Criminal Complaint Within 1 Years					
Not Prosecuted	-0.11***	-0.08***	-0.21**	-0.15*	
	(0.00)	(0.00)	(0.09)	(0.09)	
			[-0.38, -0.03]	[-0.32, 0.04]	
ADA Leniency					-0.08
					(0.05)
Observations	74072	74072	74072	74072	74072
Mean Dep Var Prosecuted	0.28				
Mean Dep Var Prosecuted Compliers	0.41				
p-value Randomization joint F-test	0.16				
Panel B: Criminal Complaint Within 3 Years					
Not Prosecuted	-0.15***	-0.11***	-0.44***	-0.39***	
	(0.01)	(0.01)	(0.10)	(0.10)	
			[-0.64, -0.25]	[-0.58, -0.19]	
ADA Leniency					-0.22***
					(0.06)
Observations	63183	63183	63183	63183	63183
Mean Dep Var Prosecuted	0.42				
Mean Dep Var Prosecuted Compliers	0.61				
p-value Randomization joint F-test	0.35				
Panel C: Criminal Complaint Within 5 Years					
Not Prosecuted	-0.15***	-0.11***	-0.39***	-0.33***	
	(0.01)	(0.01)	(0.10)	(0.10)	
			[-0.59, -0.18]	[-0.53, -0.12]	
ADA Leniency					-0.21***
					(0.07)
Observations	57230	57230	57230	57230	57230
Mean Dep Var Prosecuted	0.48				
Mean Dep Var Prosecuted Compliers	0.65				
p-value Randomization joint F-test	0.20				
Court x Time FE	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	No	Yes	No	Yes	Yes

Notes. This table reports OLS and two-stage least squares estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within different time horizons. The regressions are estimated on different samples to allow for the appropriate time to elapse to measure new complaints. The dependent variables are identified in the panel headings. Each panel reports the mean of the dependent variable for all prosecuted defendants and for prosecuted defendants within the set of compliers, as well as the p-value for an F-test of the joint significance of variables in a regression of ADA leniency on covariates, as in column 2 of Table B.1. See Appendix A.2 for details on the calculation of mean outcomes among prosecuted compliers. Two-stage least squares models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-year-month and court-by-day-of-week fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses in columns (1)-(4). For the IV estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table B.6: Probability of Subsequent Criminal Complaint Within Two Years
By Subsequent Crime Type

			Crime Type				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Any	Number	Violent	MV	Disorder/ Theft	Drugs	Other
Panel A: Any Complaint within 2 Years							
Not Prosecuted	-0.29***	-1.71**	-0.13**	-0.09	-0.19***	-0.06	0.01
	(0.10)	(0.67)	(0.06)	(0.06)	(0.07)	(0.07)	(0.04)
	[-0.49, -0.07]	[-3.15, -0.42]	[-0.26, -0.01]	[-0.21, 0.03]	[-0.33, -0.06]	[-0.19, 0.09]	[-0.07, 0.09]
Mean Dep Var Pros.	0.37	1.64	0.08	0.10	0.16	0.08	0.04
Mean Dep Var Pros. Compliers	0.55	2.84	0.20	0.15	0.23	0.11	0.02
Panel B: Misdemeanor Complaint within 2 Years							
Not Prosecuted	-0.22**	-1.04***	-0.15***	-0.09	-0.14**	-0.06	0.01
	(0.09)	(0.38)	(0.05)	(0.06)	(0.06)	(0.06)	(0.04)
	[-0.40, -0.03]	[-1.86, -0.31]	[-0.28, -0.05]	[-0.22, 0.02]	[-0.26, -0.02]	[-0.18, 0.08]	[-0.06, 0.10]
Mean Dep Var Pros.	0.24	0.95	0.05	0.09	0.12	0.06	0.04
Mean Dep Var Pros. Compliers	0.38	1.68	0.17	0.15	0.18	0.08	0.02
Panel C: Felony Complaint within 2 Years							
Not Prosecuted	-0.07	-0.53	-0.00	0.00	-0.04	-0.01	-0.02
	(0.07)	(0.34)	(0.04)	(0.01)	(0.05)	(0.04)	(0.01)
	[-0.20, 0.07]	[-1.24, 0.14]	[-0.09, 0.08]	[-0.02, 0.02]	[-0.13, 0.05]	[-0.08, 0.07]	[-0.04, 0.01]
Mean Dep Var Pros.	0.13	0.51	0.04	0.00	0.06	0.03	0.00
Mean Dep Var Pros. Compliers	0.17	0.84	0.06	-0.00	0.06	0.04	0.01
Observations	67060	67060	67060	67060	67060	67060	67060
Court x Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports OLS and two-stage least squares estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within two years, for different categories of subsequent complaints. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variables are identified in the panel headings. Each panel reports the mean of the dependent variable for all prosecuted defendants, and for prosecuted defendants within the set of compliers. See Appendix A.2 for details on the calculation of mean outcomes among prosecuted compliers. Two-stage least squares models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-year-month and court-by-day-of-week fixed effects. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses in columns (1)-(4). For the IV estimates, confidence intervals based on inversion of the Anderson-Rubin test are shown in brackets. $^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.10$.

Table B.7: Heterogeneous Effects By Demographic Group

	Male		Female		Age 18-23		Age 24-31	
	(1) OLS	(2) IV	(3) OLS	(4) IV	(5) OLS	(6) IV	(7) OLS	(8) IV
Not Prosecuted	-0.10***	-0.30**	-0.08***	-0.26*	-0.10***	-0.16	-0.10***	-0.32
	(0.01)	(0.12)	(0.01)	(0.15)	(0.01)	(0.18)	(0.01)	(0.22)
Observations	53550	53550	13383	13383	12337	12337	16413	16413
	Age 31-40		Age $\geq$ 40		Pred White		Pred Black	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Not Prosecuted	-0.09***	-0.43	-0.09***	-0.28*	-0.10***	-0.28	-0.09***	-0.23
	(0.01)	(0.27)	(0.01)	(0.16)	(0.01)	(0.17)	(0.01)	(0.14)
Observations	14493	14493	20390	20390	24376	24376	26759	26759
	Pred Hisp		White		Black		Hisp	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Not Prosecuted	-0.09***	-0.26	-0.09***	-0.13	-0.09***	-0.24	-0.09***	-0.04
	(0.01)	(0.24)	(0.01)	(0.23)	(0.01)	(0.19)	(0.01)	(0.52)
Observations	14323	14323	20447	20447	25824	25824	8943	8943
Court x Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports two-stage least squares estimates of the impact of nonprosecution on the probability of a subsequent criminal complaint within two years, for the demographic groups specified in the column headings. The models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA following the procedure described in the text. All specifications control for court-by-time fixed effects and case/defendant covariates (excluding race/ethnicity, age, and gender). Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. The groups White, Black, and Hispanic are based on race/ethnicity data as coded by SCDAO. Predicted White, Predicted Black, and Predicted Hispanic groups are based on race probability variables, described in Appendix C.1. We label a defendant as "Predicted White" if the probability estimates suggest they are most likely to be white, "Predicted Black" if they are most likely to be Black, and "Predicted Hispanic" if they are most likely to be Hispanic. ***$p < 0.01$,**$p < 0.05$,*$p < 0.10$.

Table B.8: Separating Nonprosecution and Bail Decisions

	(1) Main	(2) Control For No Bail Leniency	(3) IV No Bail Leniency	(4) Both IVs
Not Prosecuted	-0.288***	-0.288**		-0.288**
	(0.105)	(0.116)		(0.130)
No Bail			-0.226	-0.00141
			(0.169)	(0.206)
Observations	67060	67060	67060	67060
Court x Time FE	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes

Notes. This table reports 2SLS estimates that explore the role of bail requests on the probability that a defendant receives a new criminal complaint within two years post-arraignment. Column (1) reports our main estimates from Column (4) of Table 3, Panel A. Column (2) includes as a covariate a "no-bail leniency" measure based on ADAs' propensity to request bail in other defendants' cases. Column (3) uses the no-bail leniency measure as an instrument for not receiving bail. Column (4) includes both nonprosecution leniency and no-bail leniency as instruments in the same regression. All specifications control for court-by-time fixed effects and all case/defendant covariates. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. ***$p < 0.01$,**$p < 0.05$,*$p < 0.10$.

Table B.9: Robustness to Different Time FE and Instruments

	(1)	(2)			
	Main	Court x Week FE			
Panel A: Different Court x Time FE					
Not Prosecuted	-0.29***	-0.39***			
	(0.10)	(0.13)			
	[-0.49, -0.07]	[-0.70, -0.13]			
Mean Dep Var Pros Compliers	0.550	0.616			
First-Stage Coef	0.540	0.399			
First-Stage F	57.94	22.57			
Court x Time FE	Yes	Yes			
Case/Def Covariates	Yes	Yes			
	(1)	(2)	(3)	(4)	(5)
	Non-Residualized	Shrinkage	Experience	Victimless	Crime Type
Panel B: Different IV Definitions					
Not Prosecuted	-0.38**	-0.24**	-0.33***	-0.23	-0.23**
	(0.16)	(0.10)	(0.12)	(0.14)	(0.11)
	[-0.75, -0.05]	[-0.43, -0.03]	[-0.60, -0.07]	[-0.50, 0.13]	[-0.46, 0.03]
Mean Dep Var Pros Compliers	0.878	0.570	0.484	0.518	0.494
First-Stage Coef	0.162	0.792	0.403	0.375	0.342
First-Stage F	18.43	72.36	23.93	21.05	23.78
Court x Time FE	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes

Notes. This table reports two-stage least squares results of the impact of nonprosecution on the probability of a criminal complaint within 2 years, for different time FE (in Panel A) or definitions of the instrument (in Panel B). Panel A Column (1) repeats our main 2SLS results with covariates (from Table 3). In Panel A Column (2) we include court x week fixed effects instead of court x month fixed effects. In Panel B all models instrument for nonprosecution using an ADA leniency measure that is estimated using data from other cases assigned to an arraigning ADA as described in the column header. All specifications control for court-by-time fixed effects and case/defendant covariates. Robust standard errors two-way clustered at the individual and ADA level are reported in parentheses. In Panel B Column (1) we consider a version of our leave-out instrument that does not residualize out court-by-time fixed effects, and that is thus a raw measure of ADAs' leave-out nonprosecution rates. In Panel B Column (2), we use empirical Bayes shrinkage to shrink our leniency estimates towards a prior mean of 0 (see Appendix A.3 for details). Panel B Columns (3)-(5) interact our main leave-out instrument with high/low ADA experience (as measured by above- or below-median number of nonviolent misdemeanors arraigned as of the time of this case's arraignment), with an indicator for whether the crime is categorized as victimless, and with our mutually exclusive crime types. ***$p < 0.01$,**$p < 0.05$,*$p < 0.10$.

Table B.10: Sample Share by Compliance Type

	Linear			Local Linear		
	1%	1.5%	2%	1%	1.5%	2%
Compliers	0.09	0.09	0.08	0.09	0.09	0.08
Always Takers	0.30	0.31	0.31	0.31	0.31	0.32
Never Takers	0.60	0.61	0.61	0.60	0.60	0.60

Notes. This table estimates the shares of our 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 not prosecuted by the least lenient ADA; the fraction of never-takers, π_n, by the share prosecuted by the most lenient ADA; and compliers as $1 - \pi_a - \pi_n$. Least lenient ADAs are defined by being at the 1st, 1.5, or 2nd percentile of the residualized ADA leniency distribution, and most lenient are defined as being at the 99, 98.5, or 98th percentile. The first three columns use a linear specification of our first stage as in equation 3; the latter three use a local linear specification.

Table B.11: Characteristics of Marginal Defendants

	(1) $\Pr[X = x]$	(2) $\Pr[X = x \| Complier]$	(3) Ratio
Counts = 1	0.55	0.54	0.99
Counts > 1	0.45	0.42	0.94
Misd Counts = 1	0.72	0.74	1.02
Misd Counts > 1	0.28	0.22	0.78
No Serious Misd	0.53	0.57	1.06
Serious Misd	0.47	0.37	0.80
No Misd Conviction 1 Yr Prior	0.94	0.97	1.03
Misd Conviction 1 Yr Prior	0.06	0.03	0.56
No Felony Conviction 1 Yr Prior	0.97	0.98	1.01
Felony Conviction 1 Yr Prior	0.03	0.02	0.77
Not Citizen	0.24	0.10	0.41
Citizen	0.76	0.87	1.14
Any Disorderly/Theft Charge	0.33	0.39	1.18
Any Motor Vehicle Charge	0.45	0.42	0.92
Any Drug Charge	0.21	0.04	0.19
Any Other Charge	0.12	0.08	0.63
Age $\leq$ 23	0.23	0.29	1.25
Age 24-30	0.25	0.22	0.89
Age 31-40	0.22	0.18	0.83
Age $\geq$ 41	0.31	0.29	0.96
Male	0.80	0.75	0.94
Female	0.20	0.24	1.21
(Predicted) Black	0.40	0.43	1.06
(Predicted) White	0.36	0.34	0.94
(Predicted) Hispanic	0.22	0.22	1.00
(Admin) Black	0.39	0.34	0.88
(Admin) White	0.14	0.09	0.66
(Admin) Hispanic	0.31	0.28	0.90

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 ADA leniency. See Section A.2 for more detail.

Table B.12: Reweighted OLS

	(1) OLS	(2) Decile Weigts	(3) Quart x Prev. Charge Wts
Not Prosecuted	-0.097***	-0.093***	-0.091***
	(0.005)	(0.006)	(0.006)
Observations	67060	67060	67060
Court x Time FE	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes
Demographics	No	No	No
Complier Weights	No	Yes	Yes

Notes. Column (1) recreates our main OLS estimates; Columns (2)-(3) reweight those OLS estimates by splitting the sample into mutually exclusive subgroups, calculating the shares of compliers in each subgroup (as in Table B.11), and using the share of compliers relative to the share of the estimation sample in each subgroup as a weight. Column (2) splits the sample into 10 mutually exclusive subgroups based on deciles of the predicted probability of nonprosecution estimated with all available covariates. Column (3) splits the sample into 8 subgroups by quartiles of this propensity score and by whether the defendant has a previous complaint, an important source of hetereogeneity. ***$p < 0.01$,**$p < 0.05$,*$p < 0.10$.

C Missing Data

C.1 Missing Race/Ethnicity Data

We do not observe age, gender, or race/ethnicity for all defendants in our sample. Only 1.4% of cases are missing either gender or age; we exclude these cases from our analyses. However, 16.1% of cases are missing race/ethnicity information. As reported in Table C.1, race/ethnicity data are more likely to be missing for cases that were not prosecuted (27%) than prosecuted (13%), perhaps because SCDAO staff are time-constrained and may have deemed it less important to enter this information when cases were not moving forward. Moreover, the missingness of information on race/ethnicity is correlated not only with whether a case is prosecuted, but also with whether a defendant has subsequent criminal justice involvement (see Table C.1). Defendants have at least one subsequent criminal complaint within two years post-arraignment in 26% of cases that are not prosecuted and not missing race/ethnicity information, but in only 10% of cases that are not prosecuted and missing race/ethnicity, and in 40% of cases that are prosecuted and not missing race/ethnicity information, but in only 18% of cases that are prosecuted and missing race/ethnicity. This is also shown in Table C.4 Columns (1) - (3), which report regressions where the dependent variable is race missingness in the sample meeting all other sample requirements. Race missingness is correlated not only with the nonprosecution decision (and with some case and defendant characteristics), but also with defendants' subsequent criminal justice contact.

The correlation of missing data on defendant race/ethnicity and rates of subsequent criminal justice contact is likely due to the way that defendant data are entered and stored in SCDAO electronic records. When a new case is entered into SCDAO's case management system, the administrator entering the case first searches for the defendant's name in the case database. If the administrator finds the defendant's name (possibly after further narrowing his selection by date of birth, social security number, and/or address), he selects the name to start a new case record. Any defendant demographic information already entered in the database will be auto-populated in the new case record. Any missing demographic information can be filled in, and the new information will be stored as part of the defendant's record. The likelihood that a defendant has race/ethnicity information associated with his case records is thus an increasing function of the number of times he is processed through SCDAO.

Given the potential importance of race and ethnicity to current case outcomes and future criminal justice involvement, we use defendants' last names and place of residence (Suffolk County, MA) to predict the probabilities that an individual is Black, white, and/or Hispanic, using Bayesian Improved Surname Geocoding (BISG) and the Census Bureau's surname list via the R package WRU. This is the same algorithm used by the Consumer Financial Protection Bureau for race/ethnicity prediction. These probability estimates will measure actual race/ethnicity with error, but the error will not be correlated with whether or not a case was prosecuted. We label a defendant as "Predicted White" if the probability estimates suggest they are most likely to be white, "Predicted Black" if they are most likely to be Black, and "Predicted Hispanic" if they are most likely to be Hispanic.

Table C.2 reports the results of this prediction exercise. Defendants coded as Black in the SCDAO administrative data are most likely to be predicted to be Black; defendants coded as Hispanic in the administrative data are most likely to be predicted to be Hispanic; defendants coded as white in the administrative data are most likely to be predicted to be white; and defendants coded as other race in the administrative data are most likely to be predicted to be other race. Defendants missing race information in the administrative data are approximately equally likely to be predicted to be Black, Hispanic, and white.

C.2 Missing ADA Data

As noted in the text, 67% of cases meeting all other sample criteria for the 2 year post-arraignment sample are missing information on the identity of the arraigning ADA, with arraigning ADA information missing in 64% of cases that are prosecuted, and in 75% of cases that are not prosecuted. Time-constrained administrative assistants may be less likely to prioritize entering arraigning ADA information into the office's electronic case management system when a case is not moving forward. However, as reported in Table C.3, unlike the missingness of race/ethnicity information, missingness of ADA information is not also strongly correlated with defendants' subsequent criminal justice contact. This is also shown in Table C.4 Columns (4) - (6), which report regressions where the dependent variable is ADA missingness in the sample meeting all other sample requirements. While ADA missingness is correlated with the nonprosecution decision (and with some case and defendant characteristics for which we can control), there are no correlations between ADA missingness and defendants' subsequent criminal justice contact.

The relative lack of correlation between missing ADA information and outcomes is likely due to the way that ADA information is entered and stored in the SCDAO case management database. In contrast to the way that defendant race/ethnicity information is stored in SCDAO records and is auto-populated when a defendant receives a new criminal complaint, ADA information must be entered anew for each case. The missingness of ADA information is thus likely more idiosyncratic than the missingness of race/ethnicity information.

Panel C of Table 6 reports 2SLS estimates of the effect of nonprosecution on subsequent rearrest for progressively larger samples in which missing arraigning ADA information has been imputed. Our first imputed measure of ADA assignment ("Imputation 1") identifies a) court/days with only one observed arraigning ADA, where that ADA either arraigns at least two cases or arraigns one case and there is only one other case missing ADA information, and assigns that ADA to any other cases on that court/day that are missing ADA information; and b) court/weeks with only one observed ADA, where that ADA arraigns at least four cases in total on at least two days, and assigns that ADA to any other cases in that court/week that are missing ADA information. Our second imputed measure of ADA assignment ("Imputation 2") additionally identifies a) court/days with only one observed arraigning ADA, and assigns that ADA to any other cases on that court/day that are missing ADA information; and b) court/weeks with only one observed ADA, and assigns that ADA to any other cases in that court/week that are missing ADA information. Our third imputed measure of ADA assignment ("Imputation 3") additionally identifies court/days with multiple observed arraigning ADAs but with a single modal arraigning ADA, and assigns that ADA to any other cases on that court/day that are missing ADA information. Our fourth imputed measure of ADA assignment ("Imputation 4") additionally identifies court/weeks with multiple observed arraigning ADAs but with a single modal arraigning ADA, and assigns that ADA to any other cases in that court/week that are missing ADA information.

Table C.5 reports the proportions of our main estimation sample and the four imputation samples that are missing arraigning ADA information. The imbalance on missingness of arraigning ADA information for prosecuted and not prosecuted cases progressively decreases as the imputation samples grow larger. In our largest imputed sample ("Imputation 4", containing 149,185 observations or 76.4% of the sample of nonviolent misdemeanor cases meeting all other sample criteria), 24% of prosecuted cases are missing arraigning ADA information, and 22% of not prosecuted cases are missing arraigning ADA information.

In addition, there is another indication that the missing ADA data are not biasing our main results. The first-stage estimates in our main sample (not including cases that are missing ADA information) in Table 2 imply that when ADA leniency increases from 0 to 1, nonprosecution increases by 0.54. The average nonprosecution rate in the main sample is 0.203. In the sample that is missing ADA at arraignment but meeting all other criteria, the nonprosecution rate is 0.30. Taking seriously the first-stage estimates from our main sample, this would imply that ADA leniency is $0.10/0.54 = 0.18$ higher in the missing ADA sample on average. Then taking seriously the reduced form estimates from our main sample (Table 3 Panel A Column (5), coefficient of -0.16), a 0.18 increase in ADA leniency would imply criminal complaints within two years should decrease by $0.18\ x\ 0.16 = 0.0288$. In fact, the average recidivism rate in the main sample is 0.341, while in the sample missing ADA information but meeting other criteria is 0.317, which implies that the recidivism rate is 0.024 lower in the missing ADA sample, almost exactly what extrapolating the results from the main sample would imply. This further suggests that potential bias from missing ADA information is not a large issue for our estimation.

Table C.1: Missing Race

	Not Prosecuted		Prosecuted	
	(1) Missing Race	(2) Not Missing Race	(3) Missing Race	(4) Not Missing Race
Outcomes:				
Criminal Complaint Within 2 Years	0.10	0.26	0.18	0.40
Prosecution Within 2 Years	0.06	0.20	0.16	0.37
DCJIS Record Within 2 Years	0.05	0.19	0.14	0.34
Baseline:				
Number Counts	1.45	1.63	1.73	1.75
Number Misdemeanor Counts	1.10	1.16	1.28	1.38
Number of Serious Misdemeanor Counts	0.28	0.29	0.71	0.64
Misd Conviction within Past Year	0.01	0.04	0.03	0.11
Felony Conviction within Past Year	0.00	0.02	0.02	0.06
Citizen	0.96	0.81	0.94	0.71
Disorderly/Theft	0.11	0.23	0.28	0.31
Motor Vehicle	0.67	0.62	0.41	0.32
Drug	0.01	0.03	0.15	0.19
Observations	3742	10113	7176	46522
Proportion Missing Race	0.270		0.134	

Notes. This table reports summary statistics for the samples of nonviolent misdemeanor cases meeting all other sample criteria that do and do not have information on defendant race/ethnicity, as indicated by the column headers.

Table C.2: Predicted Race vs. Administrative Race Data

		Race in Admin Data				
	(1) All	(2) Black	(3) Hispanic	(4) White	(5) Other Race	(6) Race Missing
Predicted Prob. Black	0.345	0.545	0.080	0.225	0.112	0.340
Predicted Prob. Hispanic	0.256	0.131	0.778	0.186	0.065	0.272
Predicted Prob. White	0.332	0.273	0.102	0.528	0.127	0.317
Predicted Prob. Other Race	0.067	0.051	0.039	0.062	0.695	0.071
Observations	67060	25909	9056	20529	1108	10458

Notes. Race is missing in our data for 16% of observations (13% of those prosecuted and 27% of those not prosecuted). In order to keep these observations in our sample, we predicted race for all observations based on name and location (Suffolk County, MA) using Bayesian Improved Surname Geocoding (BISG), relying on the Census Bureau's surname list via the R package WRU. This is the same algorithm used by the Consumer Financial Protection Bureau for race/ethnicity prediction. The algorithm reports a predicted probability that an individual is Black, Hispanic, or white. This table compares the BISG algorithm's predicted probabilities to the information in the administrative data. We label a defendant as "Predicted White" if the probability estimates suggest they are most likely to be white, "Predicted Black" if they are most likely to be Black, and "Predicted Hispanic" if they are most likely to be Hispanic.

Table C.3: Missing ADA at Arraignment

	Not Prosecuted		Prosecuted	
	(1) Missing ADA	(2) Not Missing ADA	(3) Missing ADA	(4) Not Missing ADA
Outcomes:				
Criminal Complaint Within 2 Years	0.21	0.22	0.36	0.37
Prosecution Within 2 Years	0.15	0.16	0.32	0.34
DCJIS Record Within 2 Years	0.14	0.15	0.29	0.31
Baseline:				
Number Counts	1.62	1.58	1.71	1.75
Number Misdemeanor Counts	1.17	1.14	1.41	1.37
Number of Serious Misdemeanor Counts	0.29	0.29	0.72	0.65
Misd Conviction within Past Year	0.03	0.03	0.11	0.10
Felony Conviction within Past Year	0.01	0.01	0.06	0.05
Citizen	0.90	0.85	0.82	0.74
Disorderly/Theft	0.15	0.19	0.27	0.31
Motor Vehicle	0.64	0.63	0.44	0.33
Drug	0.02	0.03	0.13	0.18
Observations	42867	14184	97540	54748
Proportion Missing ADA	0.751		0.640	

Notes. This table reports summary statistics for the samples of nonviolent misdemeanor cases meeting all other sample criteria that do and do not have information on the identity of the arraigning ADA, as indicated by the column headers.

Table C.4: Missing Race or ADA

	Missing Race			Missing ADA		
	(1)	(2)	(3)	(4)	(5)	(6)
Not Prosecuted		0.071***	0.078***		0.119***	0.119***
		(0.004)	(0.005)		(0.002)	(0.002)
Criminal Complaint Within 2 Years			-0.082***			0.003
			(0.003)			(0.002)
Complaint 2 Years X Not Pros			-0.071***			0.004
			(0.008)			(0.004)
Number Counts	-0.009***	-0.008***	-0.007***	-0.022***	-0.020***	-0.020***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Number Misdemeanor Counts	-0.003	-0.005	-0.003	0.022***	0.020***	0.020***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Number of Serious Misdemeanor Counts	0.002	0.009***	0.004*	-0.002	0.013***	0.013***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Misd Conviction within Past Year	-0.074***	-0.069***	-0.047***	-0.001	0.010***	0.009***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Felony Conviction within Past Year	-0.040***	-0.036***	-0.025***	-0.001	0.006	0.006
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Citizen	0.129***	0.126***	0.117***	0.010***	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Disorderly/Theft	-0.022***	-0.021***	-0.017***	-0.040***	-0.032***	-0.033***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Motor Vehicle	0.016***	0.008*	0.002	0.034***	0.025***	0.025***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.002)	(0.002)
Drug	-0.035***	-0.029***	-0.027***	-0.067***	-0.048***	-0.048***
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)
Constant	0.090***	0.074***	0.109***	0.674***	0.638***	0.637***
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)
Observations	67060	67060	67060	204034	204034	204034

Notes. The dependent variable in Columns (1)-(3) is a binary variable = 1 if the observation is missing race in the administrative data, and the sample is the main sample of Table 3. The dependent variable in Columns (4)-(6) is a binary variable = 1 if the observation is missing ADA at arraignment and the sample is the sample meeting all other criteria for the main sample, as in Table 6 Panel (A) Column (3).

Table C.5: Proportion of Samples Missing ADA at Arraignment

	(1) All	(2) Prosecuted	(3) Not Prosecuted
Main Sample	0.67	0.64	0.75
Imputation 1	0.58	0.56	0.63
Imputation 2	0.41	0.40	0.45
Imputation 3	0.31	0.31	0.31
Imputation 4	0.24	0.24	0.22

Notes. This table reports the proportions of our main estimation sample and of each imputation sample (as described in the text) that are missing arraigning ADA information. See Section C.2.