DISCERNING DISCRETION?

ESTIMATING PROSECUTOR EFFECTS AT CRIMINAL SENTENCING

Emma Harrington and Hannah Shaffer

INTRODUCTION

In criminal courts, prosecutors have discretion over charging decisions and plea-deal offers, which often determine defendants' ultimate sentencing outcomes. These decisions require prosecutors to balance many, potentially competing, objectives. Sentencing laws task prosecutors with reducing criminal re-offense ("recidivism") and treating similarly situated defendants similarly ("horizontal equity"). Prosecutors may also weigh the costs of incarcerating the defendant against these legal mandates. Our project aims to evaluate whether or not different prosecutors arrive at systematically different results — in terms of incarceration, recidivism, or horizontal equity — and whether progress on one aim necessarily comes at the cost of other objectives.

This draft focuses on the trade-off between reducing incarceration and criminal re-offense. This tradeoff may seem especially stark given the mechanical effect of incarceration on re-offense — the defendant cannot re-offend while in prison. Some prosecutors, however, may be able to elide this tradeoff by selectively incarcerating those defendants who are most likely to re-offend. In this way, these prosecutors can achieve both lower recidivism and incarceration than other prosecutors. Variation across prosecutors in their degree of selective incarceration would thereby attenuate the aggregate relationship between prosecutors' recidivism and incarceration effects.

Using court records from North Carolina Superior Court, we find that prosecutors systematically vary in both their incarceration and recidivism effects. A prosecutor one standard deviation above the mean in sentencing severity imposes prison sentences that are 15% longer than the mean prosecutor (12 more days relative to a mean sentence of 83 days). A prosecutor one standard deviation above the mean in re-offense has a 6% higher rate of 3-year re-offense in their caseload (3.6pp off of a mean 3-year re-offense rate of 50%).

As expected given the incapacitation effect, we find that prosecutors who impose longer prison sentences achieve lower rates of re-offense. A prosecutor who tends to impose an additional week of incarceration will tend to have a 0.84pp lower 3-year recidivism effect. Perhaps more surprisingly, however, only 24% of the systematic variation in prosecutors' recidivism effects can be explained by the systematic variation in their incarceration effects. This means that the bulk of the systematic variation in re-offense cannot be attributed to differences in prosecutors' incarceration effects.

The variation in re-offense effects must instead be driven by differences in the degree to which prosecutors selectively incarcerate those defendants who are most likely to re-offend. Put differently, much of the variation in prosecutors' recidivism effects stem from where prosecutors lie vis-a-vis the aggregate incarceration-recidivism frontier that reflects the reduced-form tradeoff between incarceration and recidivism. In this paper, we loosely refer to prosecutor deviations from this aggregate frontier as "skill" and empirically estimate skill with the component of a prosecutor's recidivism effect that is independent of the incarceration effect. We find that the variation in skill across prosecutors is similar in magnitude to the variation in unconditional recidivism: a prosecutor one standard deviation above the mean in skill can achieve a 3.1pp lower rate of re-offense than the mean prosecutor.

This draft presents preliminary results on the variance and covariance in prosecutor skill and its relationship with horizontal equity and other dimensions of prosecutors' sentencing tendencies. In future work, we will more fully explore the drivers of these effects and their relationships. We will evaluate the relationship between prosecutor skill and other potential outcomes, such as racial equity and the incarceration of young defendants. We also plan to estimate the relationship between prosecutor skill and prosecutor tenure to learn if incarceration more closely tracks the likelihood of re-offense as prosecutors gain experience. We similarly plan to estimate the relationship between prosecutor skill, measured in the first three years, and prosecutor total tenure to learn if skill predicts prosecutor persistence on the job. In complementary work, we will leverage

North Carolina's sentencing guidelines — which discontinuously impact the degree of prosecutorial discretion over outcomes of observably similar defendants — to evaluate how prosecutor effects and their relationships change under counterfactuals with more or less discretion.

Section 1 of this paper discusses related literature. Section 2 outlines our empirical strategy, detailing how we estimate prosecutor effects and evaluate their variances and covariances in finite samples. Finally, this section compares our approach to other common estimation designs, describing the types of heterogeneity that our estimates do — and do not — capture in prosecutor effects. Section 3 introduces our empirical setting and provides summary statistics of core criminal justice outcomes. It also details the level at which we are claiming conditional random assignment of cases to prosecutors and, therefore, how we construct our set of comparisons to evaluate prosecutors. Finally, this section assesses the degree of balance in case characteristics across prosecutors. Section 4 presents and discusses the variance results for prosecutor effects on incarceration, recidivism, skill, and horizontal equity. Finally, Section 5 presents the covariances in these effects and focuses on interpreting the relationship between incarceration and recidivism effects.

1 Related literature

There is an extensive literature documenting heterogeneity in the effects and efficacy of individual decision-makers, ranging from teachers' effects on test-scores to judges' effects on sentencing. In many of these settings, it is reasonable to evaluate decision-makers along a single dimension of performance. For example, many states explicitly mandate that judges *only* consider flight risk when setting bail in criminal courts [Kleinberg et al., 2017]; auction-houses direct auctioneers to maximize their sale of cars [Lacetera et al., 2016]; firms direct hiring managers to select productive workers [Hoffman et al., 2017]; and schools direct teachers to promote student achievement [Kane and Staiger, 2008, Rothstein, 2010, Chetty et al., 2014]. Therefore, in many of these settings, it may be fair to evaluate these decision-makers according to their performance along a single dimension.

However, in many other settings, decision-makers balance multiple objectives. This complicates their evaluation, since a decision-maker's impact might advance one objective while impeding another. This paper considers prosecutors' sentencing outcomes in criminal cases, where a single, clear-cut objective is absent, and attempts to grapple with the many potential dimensions of this context.

Recently, there has been increasing attention paid to the multi-dimensional nature of some settings. For instance, Jackson [2018] documents the relationship between teacher effects on test-scores and non-test-score outcomes. Chan Jr et al. [2019] evaluate the relationship between radiologists' diagnosis rates and their diagnostic skill. And Arnold et al. [2020] estimate bail judges' effects on racial equity and the relationship between judicial discrimination and responsiveness to risk of re-offense.

Unlike the settings in these papers, there are more than two meaningful dimensions to consider at criminal sentencing. While it may be tempting to focus on incarceration only, a two-dimensional view of sentencing may miss critical aspects of prosecutors' decision-making. In addition, prosecutors with the lowest effects on recidivism and incarceration may have positive or negative effects on other potential outcomes. For instance, prosecutors who aim to treat similarly situated defendants similarly may be less likely to selectively incarcerate those defendants with the highest risk of re-offending. If horizontal equity undermined the selective incarceration of high-risk defendants, there would be direct trade-offs between efficiency and equal treatment at sentencing. Similarly, given the elevated rates of re-offense among young defendants, prosecutors who aim to give second chances to young defendants may have higher rates of re-offense than other prosecutors with similar incarceration effects. In future drafts, we hope to estimate prosecutor effects along other key dimensions — including racial equity, incarceration of young defendants, adherence to the state sentencing guidelines, and average time between first charging and final disposition — and then assess how prosecutors with low incarceration and recidivism effects stack up on these other dimensions.

2 EMPIRICAL STRATEGY

To set the stage for our empirical strategy, we consider a simple thought experiment. Suppose two defendants are arrested for the same felony offense and underlying behavior in the same community in the same time-period. After their initial appearances in court, the two cases are assigned to different prosecutors. At this point, their paths may diverge. One prosecutor may choose to extend a plea offer with supervised probation, while the other might press for incarceration. The released defendant may go on to re-offend immediately. During the time that the incarcerated defendant is in prison, he cannot re-offend in his community. However, if the incapacitated defendant had instead been assigned to the more lenient prosecutor and been released, he may or may not have re-offended. The initial assignment to a prosecutor, therefore, can affect the initial punishment as well as subsequent outcomes for defendants and their communities.

If this pattern were repeated over many defendants, the first prosecutor would have a higher rate of re-offense but lower rate of incarceration, while the second would release fewer defendants but also have fewer re-offenses. Such a difference in outcomes would be consistent with the prosecutors facing the same trade-off between incarceration and re-offense but placing different relative weights on the costs of incarceration and re-offense. Estimating the reduced-form impacts of many prosecutors would then trace out the homogeneous, aggregate possibility frontier between incarceration and re-offense. However, this simple trade-off may be complicated by the selection of *which* defendants prosecutors choose to incarcerate. If some prosecutors selectively incarcerate the defendants most likely to re-offend, these "skilled" prosecutors would attain lower rates of re-offense for a given level of incarceration. Differences in outcomes across prosecutors, therefore, may reflect differences in prosecutors' possibility frontiers rather than solely reflect differences in their preferences.

This paper aims to mirror and extend the thought experiment of the two similar defendants assigned to different prosecutors. It attempts to estimate the variation in prosecutors' causal effects on incarceration and recidivism within a given broad class of criminal offense and in a given geography and time.

The aggregate variation across prosecutor effects captures the extent to which society could affect change in a particular criminal justice outcome by changing *who* prosecutes cases. With considerable variation in prosecutor effects, it would be relatively easy to affect change by firing or retaining select prosecutors. Symmetrically, as variation in prosecutor effects approached zero, it would become impossible to change outcomes by firing or retaining select prosecutors.

We analogously construct estimates of the covariances of prosecutor effects on different criminal outcomes. The aggregate covariances capture the trade-offs that prosecutors face during plea negotiations — as well as the potential trade-offs society faces from selectively firing or retaining prosecutors based on their effect on one outcome. Due to the incapacitation effect, for instance, one might expect that prosecutors with high incarceration effects tend to have low recidivism effects — and that firing prosecutors with high incarceration rates would lead to an increase in the prevalence of re-offense. To determine whether any prosecutors can elide these tradeoffs, this paper also attempts to adjust the evaluation of a prosecutor's recidivism effect in light of the aggregate relationship between incarceration and re-offense.

We first describe our strategy for estimating the variance and covariance in prosecutor effects and then detail our methods of assessing prosecutor "skill."

A Estimating Prosecutor Effects

Our framework estimates prosecutor effects within a set of distinct "sentencing contexts." Each sentencing context is determined by the type of criminal conduct c, the geographic location of the arrest g, and the time-period t of the case. We let i index a specific case and let p denote the prosecutor assigned to the case. The case's outcome Y — e.g. incarceration — depends on the incarceration effect of the assigned prosecutor, μ_{pgct}, and the case's idiosyncratic characteristics, ϵ_{ipgct}. Thus, we have:

$$Y_{ipgct} = \mu_{pgct} + \epsilon_{ipgct}. \tag{1}$$

When we take this estimating equation to the data, we first residualize outcomes by case characteristics — i.e., a summary of the defendant's prior criminal history, the presumptive punishment according to the sentencing guidelines, defendant demographics, and the type of defense attorney (i.e. public, private, or court appointed). We include these controls for two reasons: first to account for any imbalances across prosecutors in case characteristics; and, second, to account for idiosyncratic variation in case characteristics between prosecutors in finite samples.[1] We then estimate each prosecutor's effect as the average of these residuals in her cases within a particular sentencing context. Since all of our estimates rely on comparisons *within* sentencing contexts, we always demean the prosecutor's average by the aggregate average across all the prosecutors in the sentencing context when constructing our summary measures of the variance and covariance in prosecutor effects.

Estimating and Interpreting Variances and Covariances in Prosecutor Effects. Within each sentencing context (or a crime, geography, and time period), one key estimand of interest is the variation in the prosecutors' causal effects:

$$\mathrm{Var}_{gct}(\mu_{pgct}) = \mathbb{E}\left[\sum_{pgct}(\mu_{pgct} - \bar{\mu}_{gct})^2\right]. \tag{2}$$

This expression captures the extent to which the identity of the assigned prosecutor causes outcomes to diverge from those of other prosecutors in the same sentencing context. We can then aggregate over all offices, offense-types, and time-periods to estimate the overall, average varia-

[1] As in Chetty (2014), we estimate the coefficients on these controls in a regression that includes prosecutor X office X offense X time fixed effects fixed effects to ensure that the coefficients on case characteristics are not biased by any selection of cases to prosecutors. We then use the coefficients from this first-step regression to residualize prosecutor's outcomes.

tion in estimated prosecutor effects:[2]

$$\sum_g \sum_c \sum_t \frac{N_{gct}}{N} \cdot \mathrm{Var}_{gct}(\hat{\mu}_{pgct}) = \sum_p \sum_g \sum_c \sum_t \frac{N_{pgct}}{N} (\hat{\mu}_{pgct} - \hat{\bar{\mu}}_{gct})^2$$

For the purposes of estimation, we define c to be a broad crime category such as theft or violence; we define geography g as a prosecutor's office, which often accords with a county; we define t as a 5-year time-block. Limiting the scope of our comparisons to within sentencing contexts means that we difference away any systematic differences in defendant characteristics and prosecutor practices across crime-types, places, and times. As discussed in Section B, this allows prosecutors to differ in their tendencies across time and crime. But this strategy differences away any selection of prosecutors into offices or crime-types or any real change in prosecutor effects over time.

Correcting for Sampling Error. Assuming the random assignment of cases within a given sentencing context, comparing each prosecutor's average outcomes to the average in her sentencing context will yield an unbiased assessment of her causal effect relative to other prosecutors in the same sentencing context:

$$\mathbb{E}[\bar{Y}_{pgct} - \bar{Y}_{gct}] = \mu_{pgct} - \bar{\mu}_{gct} + \underbrace{\mathbb{E}[\bar{\epsilon}_{pgct} - \bar{\epsilon}_{gct}]}_{0}.$$

However, a collection of unbiased estimates of individual prosecutor effects can still generate biased estimates of the heterogeneity in their effects. Consider, for instance, the raw variance in prosecutors' estimated incarceration effects. In any given sample of cases, some prosecutors will appear harsh or lenient simply because they happened to be assigned cases that were unobservably more or less severe than other cases in the same offense class, geography, and time-period. Thus, some of the estimated variation will reflect true heterogeneity in prosecutors' causal effects

[2] Instead, we could have first aggregated all prosecutor X offense X time fixed effects to form a single, composite prosecutor effect — and then estimated the variance in this overall prosecutor effect. With this alternative procedure, the covariances of the prosecutor X offense X time fixed effects would enter the variance of prosecutor effects. Therefore, creating one, composite measure for each prosecutor would reduce measured variation across prosecutors if there were meaningful heterogeneity in effects across offense types.

while another component will reflect each prosecutor's particular random draw of cases:

$$\mathbb{E}[\sum_{pgct}(\bar{Y}_{pgct} - \bar{Y}_{gct})^2] = \sum_{pgct}(\mu_{pgct} - \bar{\mu}_{gct})^2 + \underbrace{\mathbb{E}[\sum_{pgct}(\bar{\epsilon}_{pgct} - \bar{\epsilon}_{gct})^2]}_{\neq 0}. \tag{3}$$

There are two ways to correct for this bias from sampling error: the first adjusts the estimate analytically; the second uses a split sample. The analytical adjustment of the raw variation uses the standard error from each prosecutor estimate. Specifically, the squared standard error reflects the variance in the estimates that one would expect due to sampling error. Thus, the average squared standard error from the prosecutor estimates is an unbiased estimate of the second term in equation 3. Subtracting this from the raw variation recovers an unbiased assessment of the true variation in prosecutors' effects.

Alternatively, splitting the sample of cases in half and considering the comovement in a prosecutor's estimated effects across the two splits will adjust for the role of sampling variation. Since the random draw of cases in one sample cannot predict the draw in the other, the covariance in the prosecutor effects across these splits in the sample will isolate the first term in equation 3.

The split sample method also reveals the *stability* of prosecutor effects. If the estimates were purely capturing noise rather than something causal about the prosecutor, the average outcomes of a prosecutor in one split of cases would not predict the average outcomes of the second independent split of cases — and the covariance of the two estimates would be zero.

Correcting Covariances. When we instead estimate the covariance in prosecutors' causal effects on two different outcomes, a similar bias from sampling error arises. For instance, the prosecutor's draw of cases may bias the estimated covariance between prosecutors' incarceration and recidivism rates away from the true covariance in these causal effects. If a prosecutor happens to be assigned several cases with defendants who are unobservably less likely to re-offend, the prosecutor may choose to incarcerate few of these defendants and nonetheless have few of them re-offend, thereby biasing down the prosecutor's incarceration *and* recidivism effect. In this case, sampling error would attenuate the estimated relationship between incarceration and re-offense. As with the variance in prosecutor effects, one can correct this bias analytically or with sample-splitting. Analytically, the covariance in the errors from prosecutors' incarceration and recidivism estimates capture the magnitude of the correlation of the errors. Splitting the sample and estimating the covariance in prosecutors' incarceration and recidivism rates across splits of the data also generates an unbiased estimate of the relationship between prosecutor effects.

In this draft, we use sample-splitting to correct for sampling error in the raw variances and co-variances in prosecutor effects. This is our preferred method since it more flexibly corrects for the component of the statistics due to sampling error. However, in future drafts, we will also generate analytical estimates as robustness checks.

B Included and Excluded Heterogeneity in the Prosecutor Estimates

We flexibly allow the prosecutor's estimated effect to vary depending on the defendant's type of offense — which is our best (albeit imperfect) proxy for underlying conduct.

Decision-makers' preferences over punishments may vary considerably across offense types.[3] To the extent that prosecutor effects do meaningfully differ across offense types, estimating aggregate variation across prosecutors without conditioning on offense type would attenuate the results, since prosecutor estimates would be the aggregation of multiple, potentially conflicting tendencies. Consider, for instance, that some prosecutors may punish drug possession cases less severely relative to other prosecutors but relatively more severely on breaking and entering cases. In this case, estimated variation in prosecutor effects within drug possession and within breaking and entering would both be higher than the estimated variation of the total, composite estimate of prosecutor effects.[4]

What's more, average criminal justice punishments vary considerably across offense types. In our

[3]Indeed, Mueller-Smith [2015] finds variation in judicial tendencies across offense types, suggesting that such a concern maybe present in our context.

[4]In future work, we will explicitly assess the extent of this potential attenuation effect due to heterogeneity across offense types. We will compare our estimated variance with flexible controls for offense type to the estimated variance with aggregate controls.

setting of North Carolina Superior Court from 1995 - 2009, for instance, drug possession cases receive prison sentences of 1.09 months on average (and have a 7.7% rate of incarceration greater than 6 months), while breaking and entering cases receive prison sentences of 4.2 months on average (and have a 29.7% rate of incarceration greater than 6 months). Given the variation in average punishment across offenses, it may be more meaningful to separately evaluate variation in outcomes within a given offense category.

Our estimates relative to standard fixed effect designs: While our estimates allow individual prosecutors to vary in their own effect across offense types, this flexibility means that our estimates will necessarily exclude any systematic differences across prosecutors who tend to handle more cases in certain offense types. Put differently, by restricting our comparisons to cases that share the same offense type, our estimates do not capture the potential selection of prosecutors into offense types. If, for instance, punitive prosecutors were more likely to handle cases with drug sale offenses, our estimates would understate the heterogeneity in outcomes attributable to prosecutors since each prosecutor's drug sales effect would only be relative to other prosecutors' drug sales effects. Similarly, if certain types of prosecutors choose to work in certain offices or the average type of prosecutor changed over time, our intra-office-and-time comparisons would miss all of this heterogeneity across offices and years.

Capturing the selection of prosecutors into different offense types would require separately identifying prosecutor and offense-type effects as well as assuming a constant prosecutor effect across offense types. To accomplish this, one would normally rely on an estimating equation similar to the following:

$$Y_{ipgct} = \mu_p + \mu_{gct} + X_i + \epsilon_{ipgct}. \tag{4}$$

If punitive prosecutors tended to gravitate toward drug sales cases, this specification would capture this selection by allowing some of the harsher outcomes in drug sales cases to be attributed to the fixed effects of prosecutors who handle many drug sales cases. By contrast, our approach would attribute all of the differences in outcomes to the effect of the drug sales offense.

In theory, the estimation approach in equation 4 would perform better at capturing "real" heterogeneity if: (1) prosecutor effects were constant across offense types and (2) much of the heterogeneity across prosecutors were driven by selection into crime-types rather than variation within crime types. In practice, however, separately identifying prosecutor and offense type effects relies heavily on the set of cases that connect prosecutors who handle different crime-types. The approach in equation 4 may be biased to the extent that these pivotal connection cases are infrequent or selected.

To see how connection cases may drive estimated effects, consider the following hypothetical case assignments in an office. A prosecutor, who primarily handles drug sales cases, sees a couple of robbery cases, and has a high incarceration rate in her drug caseload. If this prosecutor happens to release the defendants in her robbery cases, but robbery has high average incarceration rates, one would infer that this prosecutor was lenient relative to prosecutors who handle many robbery cases. One would then infer that drug sales cases are more severe than robbery cases since this first prosecutor, who we have inferred is lenient, incarcerates defendants in her drug sales cases.[5] In this way, the cases that link different crime-types via a shared prosecutor are given outsized influence in equation 4. To the extent that these cases are infrequent, sampling error may significantly bias the estimated prosecutor effects. Similarly, selection in connection cases may significantly bias the estimated prosecutor effects. Our estimation approach still includes these connection cases. However, we do not attempt to use these cases to rank order prosecutors across offense types and therefore do not place *extra* weight on them.[6] [7]

[5]Further, imagine that the reverse selection also held: Robbery prosecutors saw few (and perhaps relatively minor) drug sales cases and so do not push to incarcerate their drug sales defendants, despite the overall high incarceration rate in drug sales cases. This selection of cases would turn the above set of inferences on its head – and the rank ordering of prosecutors and offenses would become still more opaque.

[6]In our approach, as prosecutors' connection cases form a smaller and smaller share of the overall caseload in a offense type, these cases will drive the resulting estimates less and less. By contrast, under the other design, the weight on connection cases will not go to zero as their frequency approaches zero. Instead, connection cases become increasingly pivotal in comparing prosecutors across different charge specialties.

[7]Our estimation strategy further mitigates the potential for connection cases to introduce bias in two other ways. First, in offices with high degrees crime-type specialization (where the potential selection of connection cases is most pronounced), we organize prosecutors into units. Each unit is treated as its own office, and prosecutors are only compared to other prosecutors in the unit. Second, to mitigate we require that a prosecutor handle at least 25 cases in a time-period X crime-type cell to enter our analysis.

In future work, we will explicitly assess the extent of selection into offense types and office locations. In one strategy, we can predict what offense types prosecutors handle and their office locations according to a measure of their baseline effect. For instance, we could estimate the selection of prosecutors into units that handle a particular offense type using prosecutors' estimated severity before they entered the unit. (This change over time could be due either to prosecutors becoming more specialized in offense type over their tenure or offices introducing units over time.) We can similarly estimate the selection of prosecutors to offices by assessing whether cross-office movers are systematically selected. An alternative strategy would use estimating strategy from equation 4, but restrict to offices that do not have units with offense type specialization.

3 DATA & BALANCE

Our empirical setting is North Carolina Superior Court, which handles most serious crimes in the state.[8] The North Carolina court system is organized into offices headed by elected District Attorneys. Each office handles most cases arising out of crimes that occur in a cluster of proximate counties as pictured in Figure 1.[9] From 1996 - 2009, the time period our data spans, there were 40 District Attorney offices. Of these, 13 had multiple physical offices spread across different counties or cities (with a maximum of four and average of 2.5).[10]

In our core sample, we only include cases involving new felony crimes — which notably excludes both probation violation cases and misdemeanor cases, since Superior Court prosecutors tend to be less involved in these proceedings.[11] We also exclude all charges that fall in offense class D or higher under the state sentencing guidelines.[12] For these severe cases — including first-degree murder and statutory rape — we believe the fact pattern specificities and other unobservable char-

[8] State courts bare the brunt of criminal caseloads, accounting for 94% of felony convictions nationally in 2006 (BJS, 2006).

[9] Some cases are handled by a District Attorney out of county or by the Attorney General's Office. Finally, many of the most severe drug trafficking cases are handled by the U.S. Attorney.

[10] These statistics reflect the office organization as of 1999, a year when many DA offices were redistricted.

[11] In probation violation cases in North Carolina, the judge and the probation officer are the key players. Similarly, in misdemeanor appeals, the District Court prosecutor or appeals unit and the presiding judges are typically the key actors.

[12] Such charges include first-degree and second-degree murder, first-degree sexual offenses, armed robbery, first-degree burglary, first-degree kidnapping, and AWDWIKISI (assault with a deadly weapon with serious injury and intent to kill). Non-routine cases also include rare sexual assault charges.

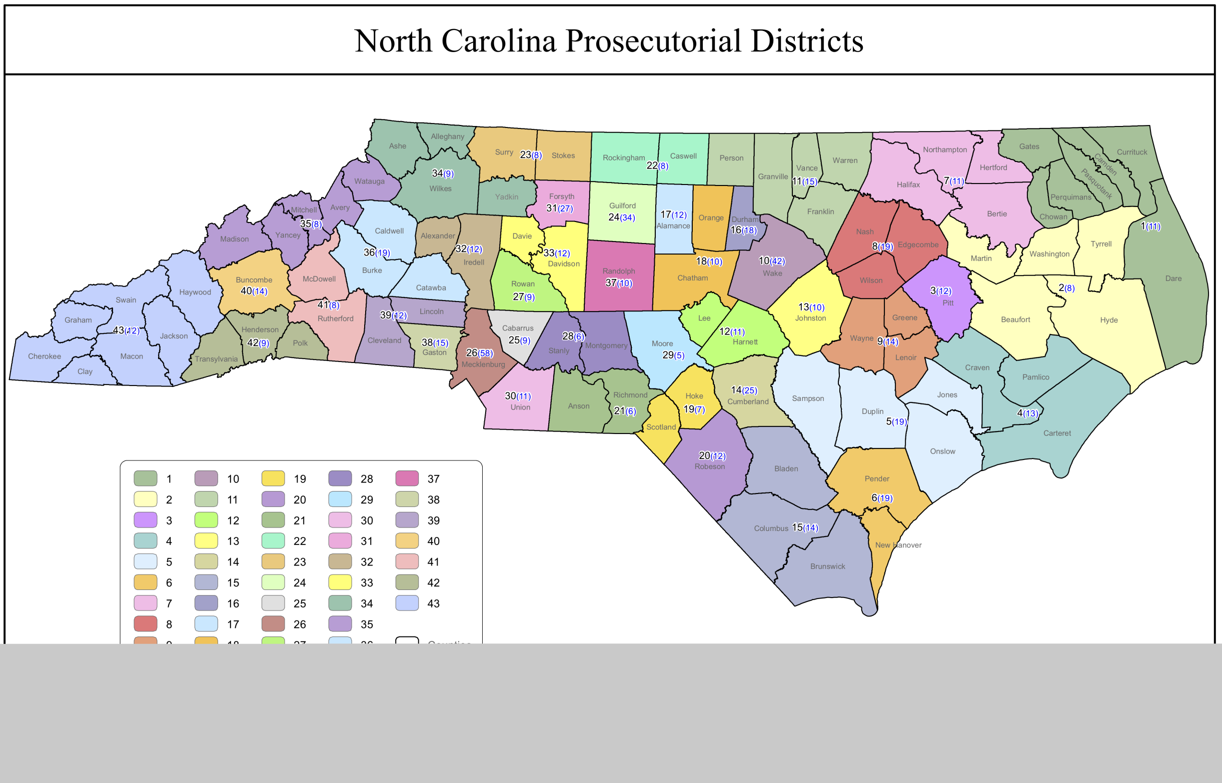


Figure 1: The current organization of District Attorney offices in North Carolina.

acteristics are more important drivers of case assignment and the ultimate sentencing outcome.

We make several other charge-related restrictions. We exclude drug trafficking cases since federal prosecutors often handle these cases, which severely biases down our recidivism measures among those defendants charged with trafficking in Superior Court. We also exclude cases with "enhancements" like Habitual Felon since these are typically charged by the prosecutor as opposed to the police. We exclude cases with lead charges that do not contain enough information to classify them into one of our offense types (e.g., Felony Conspiracy).

We require that there be at least 20 cases in a time-period X offense-type X prosecutor cell to enter our analysis, since fewer cases may violate the assumption of normality in mean errors.[13] Finally, we include only the 29 District Attorney offices that reliably record the prosecutor assigned to each

[13]We drop cases where the prosecutor is unknown or is assigned fewer than five cases in a year since these cases may reflect spurious assignments, in which the date the prosecutor was assigned to the case does not reflect the prosecutor's true dates in an office.

case.[14] Table 1 details the complete list of restrictions and their impacts on the size of the sample.

Table 1

Restriction	**Sample Size**	% **Cases Lost**
Probation Violations	930,176	39.8
1995-2009	559,935	5.55
Offices with date issues	528,882	54.43
Cases SG >= D	241,022	15.91
Misdemeanors	202,676	4.64
Sexual Offense	193,267	1.03
Drug trafficking	191,267	4.55
Habitual Felon	182,557	2.54
Non-substantative	177,920	3.01
Defendant <16	171,603	1.13
Prosecutors without names	169,662	4.87
Prosecutors, < 5 cases in year	161,402	1.61
Prosecutors, > 20 cases in Time X Offense	158,811	1.17
Cases that can be assigned to units	156,949	0.07

As shown in Figure 2, our offense categories are: drug sales, drug possession with intent to distribute, drug possession, breaking and entering and robbery, theft and fraud, reckless endangerment, assault, and sex crimes impacting children. We then estimate a separate prosecutor effect in each of these broader categories.

Table 2: Summary Statics by Crime Type

	B&E	Dr. Poss.	Dr. PWID	Dr. Sales	Negl./Reck.	Sex	Theft	Vio
Avg Caseload in FE	86.11	89.44	79.81	63.81	34.13	32.42	142.63	32.55
# Pros. in FE	346	275	267	166	99	20	420	93
% Incar >= 6mo	29.71	7.72	15.27	30.68	24.88	32.67	15.77	26.83
Avg Incar (Years)	0.35	0.09	0.16	0.4	0.33	0.65	0.18	0.52
% Recid	59.56	52.74	50.91	55.52	44.88	33.1	53.37	38.4
% Black	44.93	53.85	70.18	77.48	51.76	28.62	46.13	54.7
Avg Prior Points	2.43	1.8	1.95	2.52	2.18	0.73	2.1	1.81
Avg Def. Age	26.2	30.75	29.38	29.85	29.16	33.7	29.95	30.19

Table 2 presents summary statistics of key case outcomes by offense type, as well as the total

[14] Specifically, we eliminate offices that overwrote the initial prosecutor whenever there was a violation of probation on the original case. Since the existence of a valid prosecutor identifier was a function of the recidivism outcome of the case, these data were unusable for our analysis.

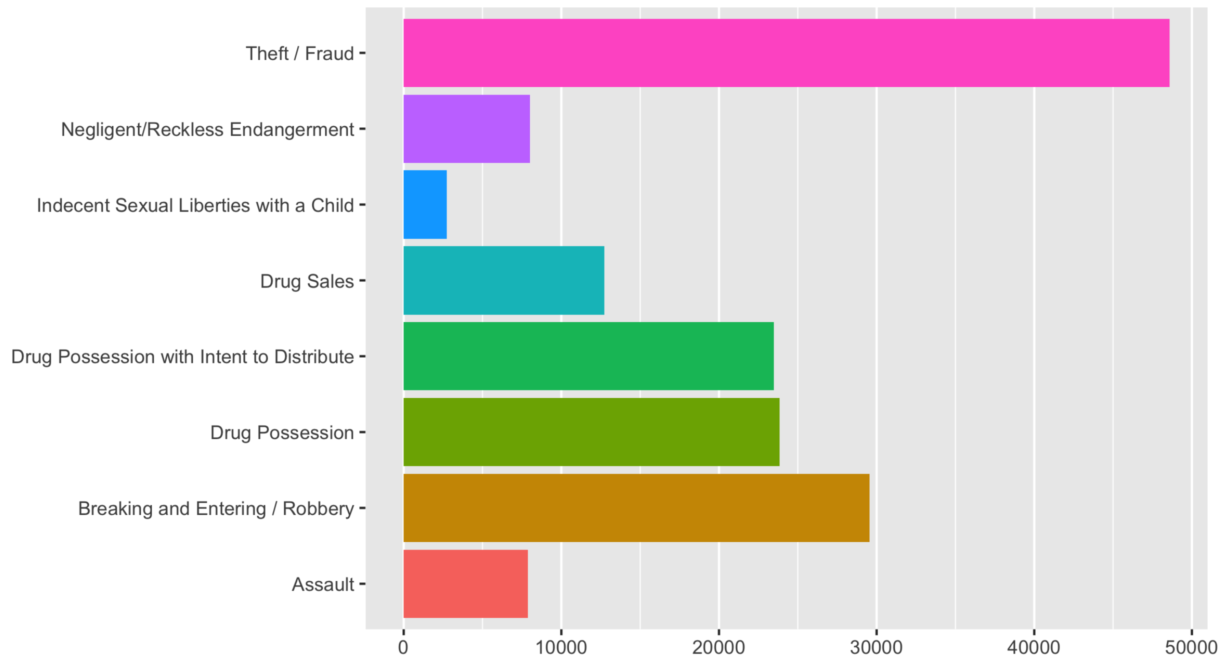


Figure 2: **Frequency of Offense Types**

number of prosecutors and average caseload of a prosecutor in a time X office X offense fixed effect within that offense type.

Balance in Baseline Characteristics: Assessing Conditional Random Assignment

Our estimates only compare prosecutors to others in the same sentencing context, in which it is plausible that cases are randomly assigned to prosecutors. Within an office and time-period, we argue that random assignment conditional on offense type is plausible because: (1) we have been told that cases are assigned based on the case's offense type; (2) case and defendant covariates are well-balanced across prosecutors in an office and time-period conditional on the case's offense type; and (3) case and defendant covariates do not meaningfully predict estimated prosecutor effects.

To estimate imbalances in case characteristics across prosecutors within an office, offense-type, and time-period, we follow the same procedure that we use to estimate prosecutor effects. The first column of Table 3 presents the standard deviation of baseline defendants characteristics across prosecutors within an office, 5-year time-period, and offense-type. These estimates isolate the

Table 3: Balance results

	Signal Std. Dev Across Prosecutors	# Additional Cases For Each 1 SD	Dependent Mean	Std. Dev Across Cases
Charges				
Avg Incar Days in 1st Charge	3.05	.	87.1	49.9
Avg Incar Days in 2nd Charge	3.47	.	33.0	44.3
% Incar in Lead Charge	0.58	0.25	19.3	10.2
Prior Criminal History				
Prior Points	0.33	.	2.0	4.0
Previously Incarcerated	2.08	0.9	15.9	36.5
Demographics				
Age at Charging	0.52	.	29.3	10.3
% Black	3.07	1.33	53.6	49.9
% Female	2.62	1.13	18.5	38.8
Predicted Outcomes				
E[Incar in Days]	5.81	.	87.1	81.3
E[Recidivate in 3 Yrs]	0.87	0.38	52.8	12.6
E[% Incarcerated]	1.39	0.6	19.3	17.5

The first column presents the standard deviation of baseline defendants characteristics across prosecutors within an office in a 5 year-block and a given crime type. These estimates isolate the component of the variation across prosecutors that is persistent across independent samples and thus likely reflect deviations from perfect random assignment. The second column translates these imbalances into cases of a prosecutor in a five-year block given the average caseload of 43. The imbalance of 3pp in the share of black defendants in a prosecutor's caseload translates into just one additional case with a black defendant over the course of 5 years within a particular crime type. Similarly, the imbalance of 1/3 of a prior point (where misdemeanors each count for 1 point and felonies count for 2 to 9 points) looks relatively small in comparison to the standard deviation in prior points of 4 in column 4.

component of the variation across prosecutors that is persistent across independent samples and thus likely reflect deviations from perfect random assignment.

For instance, a prosecutor one standard deviation above the mean in the share of black defendants has 3.07pp more black defendants than the average prosecutor. Since the average caseload for a prosecutor is 43 cases, this imbalance amounts to about 1.3 more cases with black defendants over the five-year block. Similarly, a prosecutor one standard deviation above the mean in terms of predicted incarceration[15] has cases with an average predicted prison sentence that are 5.8 days longer than those of the mean prosecutor (92.8 days off of a mean of 87 days).

The fourth column in Table 3 presents the standard deviation across all cases in each baseline characteristic, allowing us to interpret the relative magnitude of these imbalances. For instance, the standard deviation across prosecutors in the severity of a case's lead charge is about 3 days, which appears relatively small compared to the overall standard deviation of about 50 days. The next draft will report the share of the total variation in estimated prosecutor effects that can be explained by the imbalances in baseline case characteristics. This statistic will help to capture more fully the potential for these imbalances to drive our results. Future drafts will also report standard errors on the signal standard deviations in baseline case characteristics.

4 Variance Results

Variation in Incarceration and Recidivism Effects The first column of Table 4 presents the signal standard deviation across prosecutors, which is derived from the covariance between a prosecutor's effects on the same outcome estimated from two independent samples of cases. Prosecutors systematically vary in their incarceration effect. A prosecutor one standard deviation above the mean in sentencing severity imposes prison sentences that are 15% longer than the mean prosecutor (95 days relative to a mean of 83 days).

[15] We predict prison sentences in a case using defendant demographics, criminal history prior points, the offense class of the lead arrest charge, the presumptive punishment according to the sentencing guidelines, and defense attorney type.

Table 4

	Std Dev Across Prosecutors	Mean Across Cases	Total Std Dev Across Cases	Std Dev Across Prosecutors (as % of Mean)	Std Dev Across Prosecutors (as % of Total Std Dev)
3Yr Incar versus Re-Offense					
Incar Days in 3Yrs	12.1	82.8	172	14.61	7.11
% Re-Offense	3.6	52.8	49.9	6.81	7.21
E[Re-Offense | Incar] - % Re-Offense	3.14	52.8	49.9	5.94	6.29
Horizontal Equity					
E[| Y - E_p[Y|SG] |]	17.25	87.1	157	19.8	11.0
Additional Sentencing Outcomes					
Incar Days	15.38	87.1	215	17.7	7.15
% Incar	3.17	19.3	39.5	16.4	8.04
Pre-Trial Jail Days	3.7	29.0	60.0	12.8	6.18
% Felony Conviction	4.04	75.01	43.29	5.39	9.34

Variance across prosecutors: The first column presents the signal standard deviation across prosecutors derived from the covariance in prosecutor outcomes across two independent samples. The second column presents the average outcomes in cases to contextualize the magnitude of the variation across prosecutors. The third column presents the raw standard deviation in outcomes across cases, which includes both the effects of prosecutors and of case characteristics such as the arresting charge and the defendant's prior criminal history. The fourth column rescales the standard deviation across prosecutors to be as a percent of the mean across cases and the fourth column scales the standard deviation across prosecutors to be as a percent of the total standard deviation. To take an example, a prosecutor one standard deviation above the mean in terms of 3 year incarceration tends to incarcerate defendants for 12 additional days (column 1). When we compare this variation to the mean incarceration length of 82.8 days (column 2), we see that a defendant whose case is assigned to a prosecutor one standard deviation above the mean is likely to have a 14.6% increase in incarceration length (column 4). If we instead compare the variation across prosecutors to the total variation across all cases, we see that a defendant whose prosecutor is one standard deviation above the mean has an increase in incarceration length of 7% of the raw standard deviation across cases.

Perhaps more surprisingly, prosecutors also vary systematically in their recidivism effects. A prosecutor one standard deviation above the mean in re-offense has a 6% higher rate of 3-year re-offense in their caseload (3.6pp off of a mean 3-year re-offense rate of 50%). In this draft, re-offense is measured only within the three years following the final disposition date of the case. The next draft will augment these results by considering re-offense in the period 3 to 5 years following final disposition. In future work, we will also consider a more continuous measure of recidivism that integrates information about the *timing* of re-offense relative to the release date. Analyzing re-offense timing will help to reveal the prevalence of re-offense beyond the 3-year window that the current draft considers.

A Estimating Variation in Horizontal Equity

Horizontal equity — treating similarly situated defendants similarly — can be defined in many ways. In this draft, we consider horizontal equity from the perspective of the prosecutor. We evaluate the internal consistency of a prosecutor's punishments, considering how much one would expect prospecutor p's punishments to vary among observably similar defendants. For each prosecutor, we estimate the absolute gap between a defendant's sentence and the average sentence that the prosecutor imposes for that defendant type. We define defendant type as defendants arrested in the same time-period, on the same offense type and offense *class*, assigned to the same office, and with a similar criminal histories.

This definition accords with North Carolina's sentencing guidelines. Under the state's structured sentencing regime, each case has a "presumptive punishment" determined by a sentencing grid. In this regime, the defendant's current "offense class," determined by his convicted charges, and his criminal history, as summarized by his "prior points" score, jointly determine his cell in the sentencing grid. This cell prescribes a punishment type — i.e. incarceration or the option of supervised probation or incarceration — and a range for the sentence duration. [16] Under this definition, therefore, a prosecutor's effect on horizontal equity captures the average absolute gap between a defendant's punishment and the prosecutor's average punishment within the relevant cell of the state's sentencing grid.

We find that prosecutors meaningfully vary in this measure of internal consistency. A prosecutor one standard deviation above the mean has a deviation from his sentencing guideline cell average that is 17.3 days higher than the average prosecutor. This deviation is 19.8% higher than that of the mean prosecutor, who deviates by 87.1 days from his average sentence in a sentencing guidelines cell. We can also consider this measure of internal consistency from the perspective of the defendant — and ask how much uncertainty a defendant faces after his arrest about his

[16] While many states have some version of structured sentencing, North Carolina's sentencing guidelines are particularly rigidly enforced.

expected punishment given the particular prosecutor assigned to his case. From this perspective, a defendant assigned to a prosecutor one standard deviation above the mean in terms of internal inconsistency should expect that his case will deviate about 17 days more from his ex-ante expectation than had he been assigned to the average prosecutor.

5 Covariance Results

	Incarceration	% Re-Offense	E[Re-Offense | Incar] - % Re-Offense
Incarceration	1	-0.49 [R^2 =0.24]	0
% Re-Offense	-0.49 [R^2 =0.24]	1	-0.87 [R^2 =0.76]
E[Re-Offense | Incar] - % Re-Offense	0	-0.87 [R^2 =0.76]	1

Table 5: Correlation in Core 3Yr Outcomes:

Table 5 confirms our expectation that prosecutors who impose longer prison sentences achieve lower rates of re-offense: raising a prosecutor's incarceration effect by one day tends to yield a 0.44pp reduction in the prosecutor's 3-yr recidivism effect. Perhaps more surprisingly, however, only 24% of the systematic variation in prosecutors' recidivism effect can be explained by the systematic variation in their incarceration effects. This means that the bulk of the systematic variation in re-offense across prosecutors cannot be attributed to differences in their incarceration effects. The variation in re-offense effects must instead be driven by differences in the degree to which prosecutors selectively incarcerate those defendants most likely to re-offend, thereby beating the aggregate relationship between incarceration and re-offense.

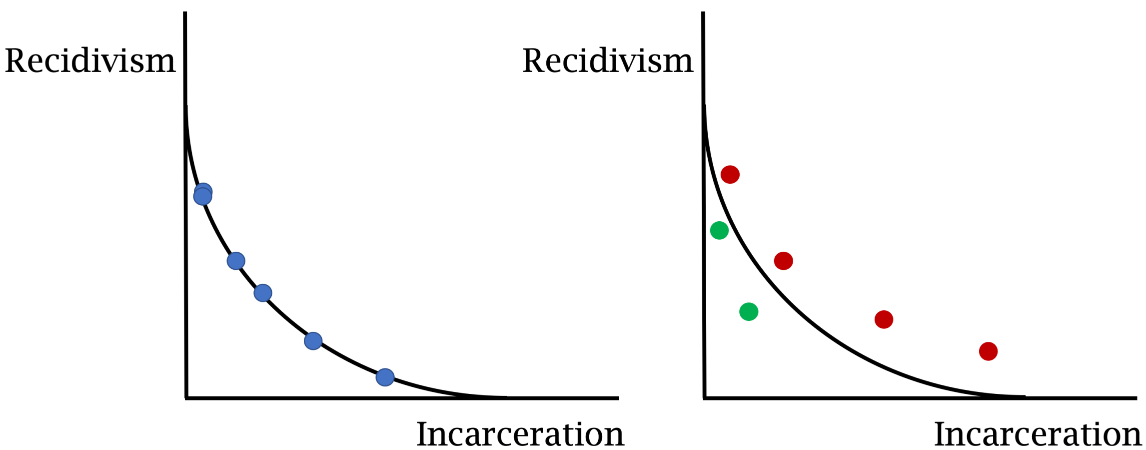
A Decomposing the Covariance between Incarceration and Recidivism: Estimating Prosecutor Skill

This subsection considers the drivers of the aggregate possibility frontier of incarceration and recidivism. Variation in recidivism effects and the covariance of incarceration and recidivism effects may be driven by two sources of heterogeneity across prosecutors: (1) preferences — the relative weights prosecutors place on the costs of incarceration and re-offense — and (2) information about defendant recidivism risk. In this simple decomposition, therefore, each prosecutor's recidivism effect is a function of: (1) their incarceration rate and (2) the selection of defendants chosen for release.

If prosecutors vary in their preferences over incarceration and recidivism but have identical information about defendant risk of re-offense, they will choose to locate on different points of the same incarceration–recidivism frontier. Since incarceration mechanically reduces recidivism by incapacitating defendants, a prosecutor who prioritizes reducing recidivism will tend to have higher incarceration and lower recidivism effects, all else equal. In Figure 3(a), such a prosecutor might be in the lower right extreme of the possibility frontier shared by all prosecutors.

Assuming homogeneous information, the aggregate estimated covariance would fully capture the trade-off that each prosecutor faces between incarceration and recidivism. However, variation in prosecutor "skill" elides such a rigid trade-off. Heterogeneity in skill would enable certain prosecutors to lie on distinct possibility frontiers. All else equal, prosecutors who selectively incarcerate only those defendants with the highest recidivism probabilities can achieve lower incarceration rates for a given rate of recidivism. The outcomes of these prosecutors would lie below (or closer to the origin in) the frontier that averaged across all prosecutors in Figure 3(b). To estimate the heterogeneity in prosecutor "skill," we evaluate whether a prosecutor's incarceration and recidivism outcomes systematically differ from the aggregate frontier. [17] As a first pass in estimating prosecutor skill, we consider prosecutor deviations from an estimated linear frontier between incarceration and recidivism. Specifically, we use the gap between a prosecutors realized recidivism estimate and the expected recidivism effect, given that prosecutor's incarceration estimate and the aggregate relationship between prosecutor incarceration and recidivism effects. Suppressing

[17] Of course, prosecutors with preferences over observable or unobservable case characteristics that are correlated with recidivism risk would impact these estimates. Our estimate of prosecutor skill simply captures the selection of released defendants in terms of recidivism risk, rather than any recidivism risk motivation and knowledge of the prosecutor.



(a) Homogeneous Skill (b) Heterogeneous Skill

Figure 3: If prosecutors do not meaningfully vary in their "skill" — the degree to which they selectivity incarcerate those defendants most likely to re-offend, as illustrated in (a) — then all prosecutors' outcomes will lie on the same possibility frontier for recidivism and incarceration. If, instead, as (b) illustrates, prosecutors differ in the degree to which incarceration outcomes align with recidivism risk, prosecutors may attain recidivism and incarceration estimates that do not lie on the aggregate frontier.

notation for office, time, and offense:

$$\text{Skill}_p = \hat{\mu}_{\text{Recid},p} - \frac{\hat{\text{Cov}}(\mu_{\text{Recid},p}, \mu_{\text{Incar},p})}{\hat{\text{Var}}(\mu_{\text{Incar},p})} \hat{\mu}_{\text{Incar},p} \tag{5}$$

However, we need to assess whether the observed deviations of prosecutors' outcomes from the aggregate frontier are larger than what chance alone could explain. In a given sample, a lucky draw of defendants with low likelihoods of re-offense would create the illusion that some prosecutors can achieve lower rates of re-offense at a given level of incarceration. Sampling error would therefore create a picture the looked like Figure 3(b) even if the world were closer to the picture in Figure 3(a). If deviations from the aggregate frontier were due solely to the luck of a particular draw of cases, prosecutor deviations would not replicate across a new sample of cases. But if some of the variation were due to skill, estimates of prosecutor skill would persist across samples. Therefore, to isolate the variation due to the systematic component of prosecutors' skill, we split our data in half, estimate prosecutor skill in each split, and then estimate their covariance.

Prosecutor Skill Results

Table 4 shows that the variation in skill across prosecutors is similar in magnitude to the variation in unconditional recidivism: a prosecutor one standard deviation above the mean in skill can achieve a 3.1pp lower rate of re-offense than the mean prosecutor.

However, this estimated variance requires some qualification. As with sampling error, diminishing returns of incarceration in reducing recidivism may also generate a picture that looks like Figure 3(b), even if the world were closer to the picture of Figure 3(a). If all prosecutors were equally likely to incarcerate defendants with higher risks of re-offense, those prosecutors with low incarceration effects may have an easier time selecting the very riskiest defendants to incarcerate. By contrast, prosecutors with high incarceration effects may have already incarcerated the most risky defendants. Put differently, the marginal defendants are likely of lower risk for high incarceration prosecutors. In a world with diminishing returns in incarceration, lenient prosecutors are more likely to appear skilled than harsh prosecutors. Consequently, there will be predictable deviations from the aggregate frontier even without heterogeneous skill. With diminishing returns, systematic deviations from the aggregate frontier offer evidence that prosecutors respond to risk of re-offense. Yet, they do not provide conclusive evidence of heterogeneity in skill. In future work, we will estimate a more flexible frontier between recidivism and incarceration.

	Correlation with Skill
% Re-Offend in 3Yrs	-0.87
% Incar > 6mos	0.28
Pre-Trial Jail Days	0.38
$E[\mid Y - E_p[Y \mid SG] \mid]$	-0.17

Table 6: Correlation in Prosecutor Effects:

The relationship between skill and severity. Table 6 reveals that the covariance between a prosecutor's estimated skill and her propensity to incarcerate defendants for more than 6 months is positive. Given that the skill estimates remove the component of recidivism that is expected given a prosecutor's incarceration effect, some explanation of this positive comovement is warranted. Skill controls for incarceration *length*, which means that the covariance of prosecutor skill and incarceration rates will only pick up the component of a prosecutor's incarceration rate that is independent of incarceration length. Among prosecutors that share the same average incarceration length, those who incarcerate more defendants must necessarily impose shorter sentences in order to achieve incarceration lengths equal to those prosecutors with lower incarceration rates. Thus, a higher incarceration rate at a given average incarceration length is associated with a compressed distribution of prison sentences.

Therefore, we interpret the positive relationship between a prosecutor's skill and 6-month incarceration rate as evidence that prosecutors with more compressed distributions of punishments achieve lower rates of re-offense. There are several possible explanations for this pattern: (1) Diminishing returns to incarceration at the defendant level; (2) correlation in preferences over punishment types with skill. The reduction in risk of re-offense from increasing the length of the prison sentence for defendants with long prison sentences may be smaller than for defendants with shorter prison sentences. This may be due to the phenomenon of "age out" — as people age, they are less likely to commit criminal (especially violent criminal) offenses. This diminishing returns to incarceration may also stem from heterogeneous treatment effects of prison across lengths of sentences. In future work, we hope to explore this relationship further.

The relationship between horizontal equity and skill In theory, one might expect a trade-off between a prosecutor's horizontal equity and her skill. Prosecutors who carefully attempt to treat similarly situated defendants similarly may be unable to act on their perceived differences in the likelihood of reoffense. However, Table 6 presents a negative correlation between equity and skill, suggesting that this trade-off may not be present in our setting.

REFERENCES

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

D. C. Chan Jr, M. Gentzkow, and C. Yu. Selection with variation in diagnostic skill: Evidence from radiologists. Technical report, National Bureau of Economic Research, 2019.

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

M. Hoffman, L. B. Kahn, and D. Li. Discretion in hiring. *The Quarterly Journal of Economics*, 133(2): 765–800, 2017.

C. K. Jackson. What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, 126(5):2072–2107, 2018.

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

J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. Human decisions and machine predictions. *The quarterly journal of economics*, 133(1):237–293, 2017.

N. Lacetera, B. J. Larsen, D. G. Pope, and J. R. Sydnor. Bid takers or market makers? the effect of auctioneers on auction outcome. *American Economic Journal: Microeconomics*, 8(4):195–229, 2016.

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

J. Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):175–214, 2010.