

DISCERNING DISCRETION: ESTIMATING PROSECUTOR EFFECTS AT CRIMINAL SENTENCING

INTRODUCTION

In criminal courts, prosecutors have discretion over charging decisions and plea-deal offers, which often determine 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 the extent to which different prosecutors arrive at systematically different results — in terms of incarceration, recidivism, or horizontal equity — and whether progress on one aim 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 — i.e., a 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 lower recidivism *and* lower 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 mechanical effect of incarceration on re-offense, prosecutors who impose

longer prison sentences achieve lower rates of re-offense. A prosecutor who imposes an additional week of incarceration tends to have a 0.84pp lower 3-year recidivism effect. These differences may reflect differences in prosecutors' priorities, or specifically, the relative weight they place on the costs of incarceration versus re-offense. If differences in priorities were the sole driver of differences in outcomes, all prosecutors would lie on the same frontier between incarceration and re-offense. And the variation in prosecutors' incarceration effects would trace out this frontier, fully explaining any variation in prosecutors' re-offense effects. Yet, variation in prosecutors' incarceration effects explains only 24% of the variation in prosecutors' recidivism effects, suggesting that differences in priorities do not drive most of the variation in re-offense effects.

The bulk of the systematic variation in re-offense 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 stems from where prosecutors lie vis-a-vis the aggregate incarceration-recidivism frontier. 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 her 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 in prosecutor skill and its covariance with horizontal equity and other dimensions of prosecutors' sentencing effects, including aggregate racial gaps and age gaps at sentencing. In future work, we will more fully explore the drivers of these effects and their relationships. We will decompose prosecutor skill into an observable component — a prosecutor's responsiveness to predictable risk of re-offense (estimated using observable, baseline case characteristics) — and an unobservable component — the difference between a prosecutor's reduced-form skill and their responsiveness to the correlates of risk. 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.¹²

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. 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 bail determinations. 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 courts³; auction-houses direct auctioneers to maximize their sale of cars⁴; firms direct hiring managers to select productive workers⁵; and schools direct teachers to

¹We 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.

²In complementary work, we also 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.

³J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237-93, 2017

⁴N. 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):195-29, 2016.

⁵M. Hoffman, L. B. Kahn, and D. Li. Discretion in hiring. *The Quarterly Journal of Economics*, 133(2) 765-800, 2017.

promote student achievement.^{6,7,8} Therefore, in many of these settings, it may be fair to evaluate 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 certain settings. For instance, Jackson (2018) documents the relationship between teacher effects on test-scores and non-test-score outcomes.⁹ Chan et al. (2019) evaluate the relationship between radiologists' diagnosis rates and diagnostic skill.¹⁰ 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.¹¹

Unlike the settings in these papers, at criminal sentencing, there are many more than two meaningful dimensions to consider. While it may be tempting to focus on incarceration and recidivism only, a two-dimensional view of sentencing may miss critical aspects of prosecutors' decision-making. And 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 incar-

⁶T. J. Kane and D. O. Staiger. Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research, 2008.

⁷J. Rothstein. Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, 125(1):175-214, 2010.

⁸R. Chetty, J. N. Friedman, and J. E. Rockoff. Measuring the impacts of teachers: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593-2632, 2014.

⁹C. 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.

¹⁰D. 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.

¹¹D. Arnold, W. S. Dobbie, and P. Hull. Measuring racial discrimination in bail decisions. Technical report, National Bureau of Economic Research, 2020.

ceration 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.

2 EMPIRICAL STRATEGY

To fix ideas for our empirical strategy, consider a simple thought experiment. Two defendants are arrested for the same felony offense and underlying behavior in the same community and time-period. After their initial appearances in court, the two cases are assigned to different prosecutors. At this point, the defendants' paths may diverge. One prosecutor may choose to extend a plea offer with supervised probation, while the other might press for incarceration. The first defendant who is released may go on to re-offend immediately. By contrast, the second defendant who is incarcerated cannot re-offend in his community during the time that he is in prison. 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 a defendant and his community.

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 prosecutor 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 and choosing the place different relative weights on the costs of incarceration and re-offense. Estimating prosecutors' reduced-form effects 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 dif-

ferences 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.

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

A Estimating Prosecutor Effects

Our framework estimates prosecutor effects within a set of distinct "sentencing strata." Each sentencing strata is determined by the type of criminal conduct c , the geographic location of the arrest g , and the time-period t of the case.¹² 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} , only some of which may be observable. Thus, we have:

$$Y_{ipgct} = \mu_{pgct} + \epsilon_{ipgct}. \quad (1)$$

When we take this estimating equation to the data, we first residualize outcomes by observable case characteristics, X_i .¹³ 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.¹⁴

¹²For 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.

¹³Specifically, we include 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).

¹⁴As 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. Finally, we estimate each prosecutor's effect as the average of these residuals in her cases within a particular sentencing strata.

When evaluating a prosecutor, we always limit comparisons to within a given sentencing strata. Thus, the core building blocks of our estimands are the prosecutor's average outcomes and aggregate average outcomes in a sentencing strata:

$$\mu_{pgct} - \bar{\mu}_{gct}$$

where $\bar{\mu}_{gct}$ denotes the average causal effect of prosecutors in a sentencing strata. Thus, in our estimating equations, any systematic differences in defendant characteristics across place, time, or crime-type are naturally differenced away. These building blocks underpin the variation in the prosecutors' causal effects within each sentencing strata:

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

We can then aggregate over all offices, crime-types, and time-periods to estimate the average variation in estimated prosecutor effects within narrow sentencing strata:¹⁵

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

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 strata, averaged over all sentencing strata in the state. This variation captures the extent to which society could affect change in a particular criminal justice outcome by changing *who* prosecutes cases, within the bounds of each sentencing context. 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

¹⁵Instead, 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.

firing or retaining select prosecutors.¹⁶

Another key estimand of interest is the comovement in prosecutor effects on different criminal-justice outcomes:

$$\text{Cov}(\mu_{pgct}^Y, \mu_{pgct}^{\tilde{Y}} | gct) = \sum_p \sum_g \sum_c \sum_t \frac{N_{pgct}}{N} (\mu_{pgct}^Y - \bar{\mu}_{gct}^Y) (\mu_{pgct}^{\tilde{Y}} - \bar{\mu}_{gct}^{\tilde{Y}})$$

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. For instance, due to the incapacitation effect, one might expect that prosecutors with high incarceration effects will tend to have low recidivism effects — and therefore that firing prosecutors with high incarceration rates would lead to an increase in the prevalence of re-offense.

Included and Excluded Heterogeneity in Prosecutor Effects. When we estimate the variation across prosecutors within a sentencing strata, we allow for full flexibility in a prosecutor’s effect across sentencing strata. We allow for this flexibility since, in practice, decision-makers’ preferences over punishments may vary considerably across offense types.¹⁷ To the extent that prosecutor effects do meaningfully differ in their effects across offense types, suppressing this variation would attenuate the within-strata variation. Consider, for instance, that some prosecutors may punish less severely than other prosecutors in drug possession cases but relatively more severely in breaking and entering cases. In this case, forcing the prosecutor to have a single effect in drug possession and breaking and entering would make the prosecutor look typical in her handling of both types of cases and mute the estimated variation across prosecutors in both strata.¹⁸

¹⁶By allowing a prosecutor’s effect to vary across time, place, and crime-type, we preclude identifying selection of prosecutors into offices or crime-types or systematic changes in prosecutor effects over time.

¹⁷Indeed, Mueller-Smith (2015) finds variation in judicial tendencies across offense types, suggesting that such a concern maybe present in our context. M. Mueller-Smith. The criminal and labor market impacts of incarceration. *Unpublished Working Paper*, 18, 2015.

¹⁸In future work, we will explicitly assess the extent of this potential attenuation effect due to heterogeneity across crime types. Among prosecutors who do not have crime-type specializations, we can directly estimate the extent of heterogeneity in prosecutor effects across crime types. We can also gain indirect evidence on heterogeneity across crime type by estimating prosecutor effects *without* limiting comparisons to being within crime types. A much lower variance across prosecutors in this specification would suggest that prosecutors do vary across crime types in their

What's more, average criminal justice punishments vary considerably across offense types. In our 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 – and the likely variation in society's preferences for punishment across offenses — it may be more meaningful to separately evaluate variation in outcomes within a given offense category.

Variation in prosecutor effects within narrow sentencing strata may understate the variation within a broader sentencing context since variation within sentencing strata differences away any systematic selection of prosecutors into crime-types or offices as well as any systematic change in prosecutor effects over time. Broadening the scope of our analysis requires stronger identifying assumptions about the homogeneity of prosecutor effects and relies more heavily on the set of "movers" who prosecute cases in different crime-types, in different years, and across different offices. In future work, we will use fixed effects designs to estimate heterogeneity in prosecutors across broader sentencing contexts. We will also directly estimate the selection of prosecutors into crime-types and offices based on baseline effects. More details on this strategy and a comparison of our approach to a standard fixed effects design can be found in Appendix 6.

B Correcting for Sampling Error

When we estimate the variation in prosecutors' causal effects – e.g., on incarceration — we must account for the effect of sampling error on our estimates. 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, one component of the estimated variation will be the "signal variance" — which reflects the true heterogeneity in prosecutors' causal effects — and another component will

effects.

be the "noise variance" — which reflects each prosecutor's particular random draw of cases :

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

We can isolate the systematic variation or "signal variance" in prosecutors' effects by splitting the sample of cases in half and considering the comovement in a prosecutor's estimated effects across the two splits. 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 6. To form a confidence interval for the signal standard deviation, we bootstrap this split sample procedure. Appendix 7 describes an alternative analytical adjustment that uses the full sample.

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 in the second independent split of cases — and the resulting covariance of the two estimates would be zero.

Correcting Covariances. When we estimate the covariance in prosecutor effects — on e.g., incarceration and re-offense — sampling error can again bias our inferences. 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. Splitting the sample and estimating the covariance in prosecutors' incarceration and recidivism rates across splits of the data generates an unbiased estimate of the relationship between prosecutor effects.

C Measuring Skill

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 1(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 1(b). To estimate the heterogeneity in prosecutor "skill," we evaluate whether a prosecutor's incarceration and recidivism outcomes systematically differ from the aggregate frontier.¹⁹ 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

¹⁹Of 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.

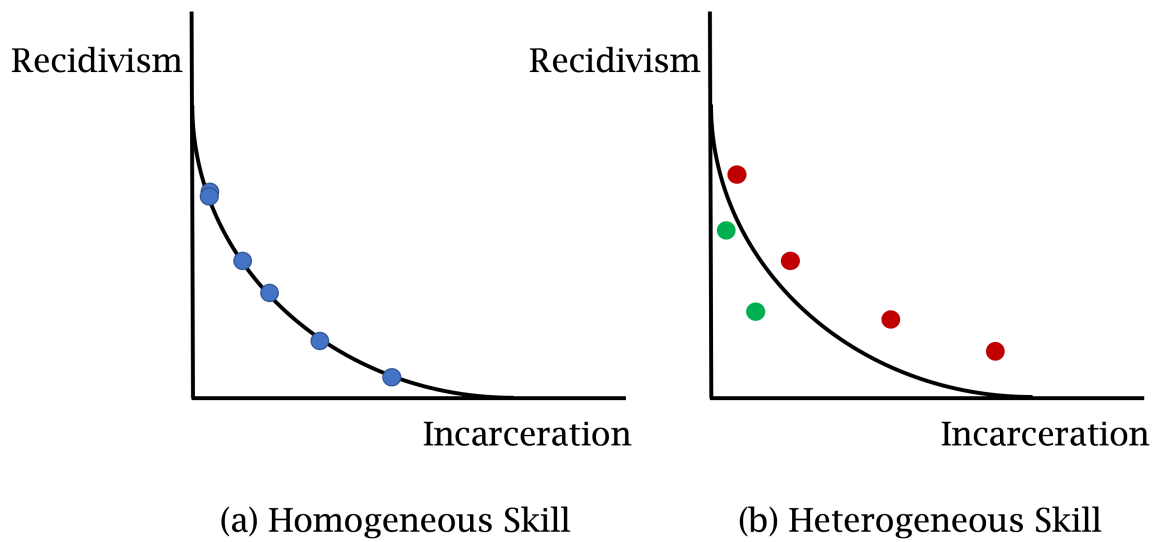


Figure 1: If prosecutors do not meaningfully vary in their "skill" — the degree to which they selectively incarcerate those defendants most likely to re-offend, as illustrated in (a) — then all prosecutors' outcomes will lie on the same possibility frontier for recidivism and incarceration. If, instead, 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. This could either be due to differences in prosecutors' information about the likelihood of re-offense or preferences about other defendant characteristics that are correlated risk of re-offense.

the aggregate relationship between prosecutor incarceration and recidivism effects. Suppressing notation for office, time, and offense:

$$\text{Skill}_p = \hat{\mu}_{\text{Recid},p} - \frac{\hat{\text{Cov}}(\mu_{\text{Recid},p}, \mu_{\text{Incar},p})}{\hat{\text{Var}}(\mu_{\text{Incar},p})} \hat{\mu}_{\text{Incar},p} \quad (4)$$

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 that looked like Figure 1(b) even if the world were closer to the picture in Figure 1(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.

3 DATA & BALANCE

Our empirical setting is North Carolina Superior Court, which handles most serious crimes in the state.²⁰ 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 1.²¹ 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).²²

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

²⁰State courts bare the brunt of criminal caseloads, accounting for 94% of felony convictions nationally in 2006 (BJS, 2006).

²¹Some 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.

²²These statistics reflect the office organization as of 1999, a year when many DA offices were redistricted.

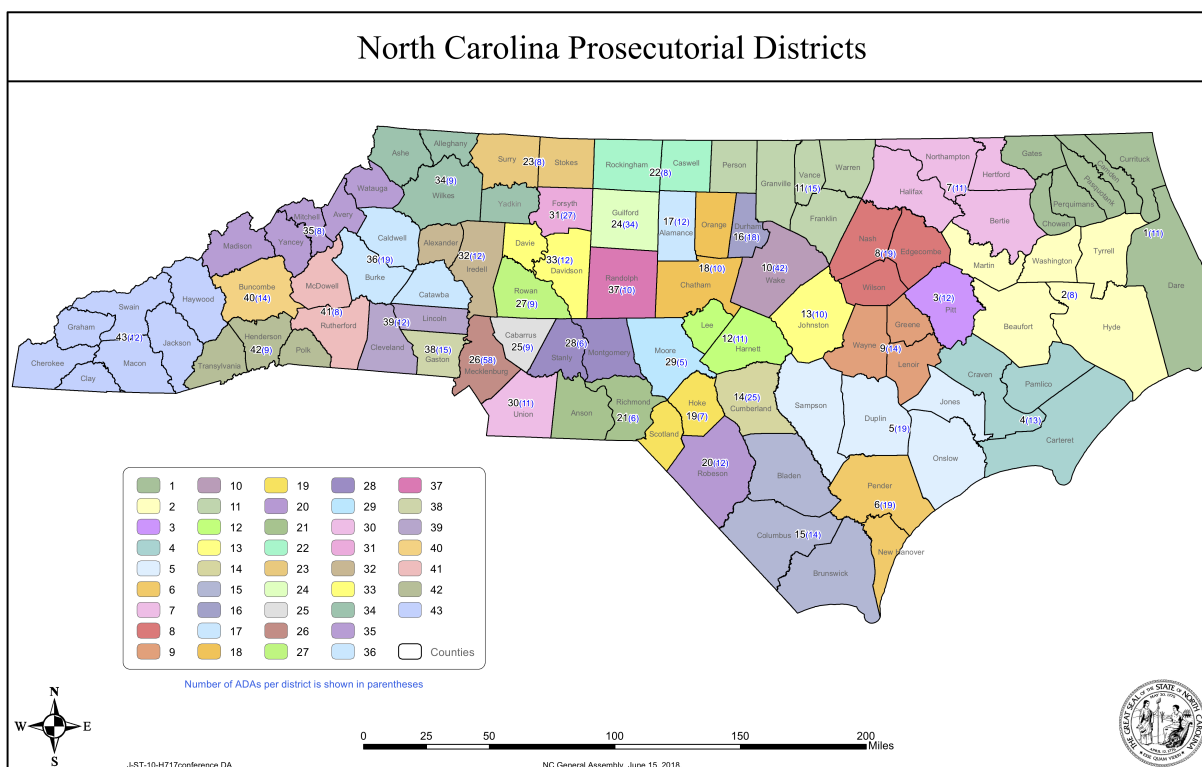


Figure 2: The current organization of District Attorney offices in North Carolina.

be less involved in these proceedings.²³ We also exclude all charges that fall in offense class D or higher under the state sentencing guidelines.²⁴ For these severe cases — including first-degree murder and statutory rape — we believe the fact pattern specificities and other unobservable characteristics 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 we organize cases into sentencing strata based on the *arresting* charge and these enhancement are typically charged by the prosecutor as opposed to the

²³In 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.

²⁴Such 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.

police. We exclude cases with lead charges that do not reflect substantive crimes — e.g. Felony Conspiracy — and thus do not contain enough information to classify them into one of our offense types.

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.²⁵ Finally, we include only the 29 District Attorney offices that reliably record the prosecutor assigned to each case.²⁶ Table 1 details the complete list of restrictions and their impacts on the size of the sample.

Table 1: Sample Restrictions

Restriction	Sample Size Before Restriction	% Cases Lost
Probation Violations	930,176	39.8
Cases assigned to prosecutors before 1995 or after 2009	559,935	5.55
Offices with data issues in recording prosecutor assignment	528,882	54.43
Charge-related restrictions		
Cases $SG \geq D$	241,022	15.91
Misdemeanors	202,676	4.64
Sexual offenses	193,267	1.03
Drug trafficking	191,267	4.55
Habitual felon	182,557	2.54
Non-substantative charges	177,920	3.01
Defendant <16	171,603	1.13
Prosecutor-related restrictions		
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 cannot be assigned to units	156,949	0.07

As shown in Figure 3, our offense categories are: drug sales, drug possession with intent to distribute, drug possession, breaking and entering and robbery, theft and fraud, reckless endanger-

²⁵We 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.

²⁶Specifically, 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.

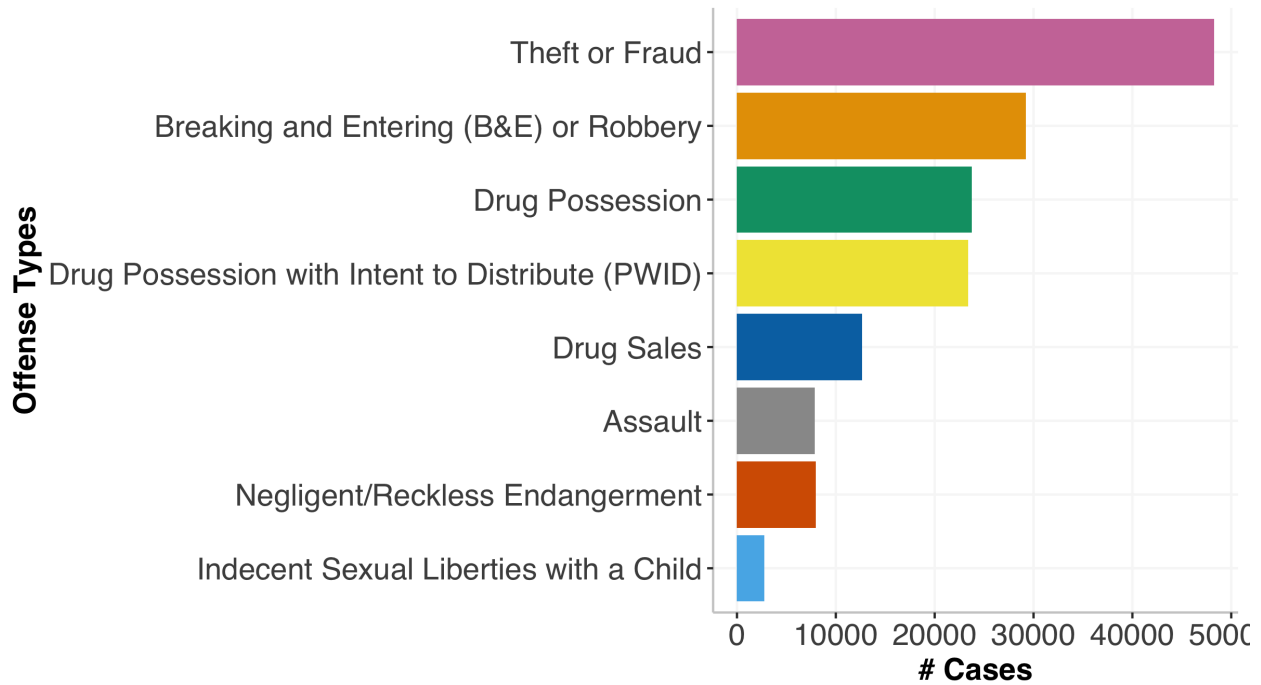


Figure 3: Frequency of Offense Types

ment, 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

	All	Theft or Fraud	B&E or Robbery	Drug Poss.	Drug PWID	Drug Sales	Assault	Reckless Endanger.	Indecent Liberties
Prosecutor Statistics									
Caseload in Crime x Time	84.1	133.9	76.4	77.1	69.2	47.8	19.5	20.5	13.5
# Pros. in Office in Crime x Time	10.8	9.9	9.6	9	9.5	8.5	9.1	9.4	5.7
# Pros. in Crime	458	415	334	268	263	162	89	93	17
Sentencing Outcomes									
% Incarcerated >= 6mos	19.3	15.4	29.2	7.5	15	30.4	26.7	24.7	32.6
Incarceration (in Days)	87.1	62.5	123.2	32.6	55.9	142.6	187.6	119.2	235.5
% Re-offend in 3Yrs	52.8	53.5	59.8	52.8	50.9	55.6	38.5	44.9	33.1
Case Characteristics									
Prior Points	2	2	2.3	1.7	1.9	2.5	1.8	2.1	0.7
% Previously Incarcerated	23.2	22.9	27.7	19.3	21.4	27.7	20.6	25.1	10.4
% Black	53.6	46.1	44.8	53.7	70.1	77.4	54.7	51.7	28.6
Age at Charge	29.3	29.9	26.1	30.7	29.4	29.8	30.2	29.1	33.7

Table 2 presents summary statistics of key case outcomes by offense type, as well as the total number of prosecutors and average caseload of a prosecutor in a time X office X offense fixed effect within that offense type.



Figure 4: This figure depicts the signal standard deviations of baseline defendant characteristics across prosecutors within an office, time-period, and offense type. We construct the confidence intervals by bootstrapping our split-sample procedure.

Balance in Baseline Case Characteristics: Assessing conditional random assignment

Our estimates only compare prosecutors to others in the same sentencing strata, 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 across prosecutors in baseline case characteristics, we follow the same procedure used to estimate prosecutor effects. Figure 4 shows signal standard deviation of baseline defendant characteristics across prosecutors within an office, time-period, and offense type. These estimates isolate the component of the variation across prosecutors that persists across independent splits of the data — and so reflects true deviations from perfect random assignment within

sentencing strata. The top two bars reflect the imbalances in the severity of the charge (i.e. the average incarceration associated with a charge) in a prosecutor's caseload. A prosecutor one standard deviation above the mean in lead charge severity has a caseload with about a 3 day higher average incarceration length.

The imbalances in prior convictions and core demographics are also positive but not large in magnitude. 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 a five-year block. And a prosecutor one standard deviation above the mean in prior conviction "points" has half an extra point on average.

The bottom two bars show the imbalances for predicted recidivism and incarceration, which are constructed using baseline case characteristics²⁷ and provide summary measures of our remaining imbalances. A prosecutor one standard deviation above the mean in terms of her caseload's predicted incarceration rate has cases with an average predicted prison likelihood that are about 10 days longer than those of the average prosecutor.

Figure 5 presents the share of the total variation in estimated incarceration effects that can be explained by the imbalances in baseline case characteristics. This statistic helps to capture the potential for remaining imbalances in case characteristics to drive our results. To construct the confidence intervals for these correlations, we split our sample in half, estimate the relationship between incarceration effects in one split and imbalances in baseline characteristics in the other split, and then bootstrap this procedure. All of the confidence intervals encompass zero — and we can rule out that more than 2.9% of the variation in our estimated incarceration effects are driven by the any one of these imbalances.

²⁷This specification includes information about charges, priors, and core demographic characteristics.

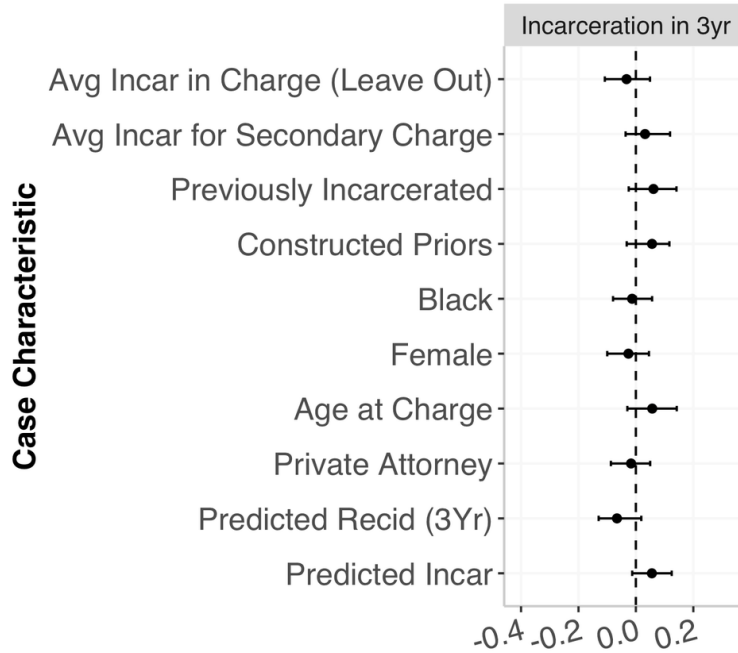


Figure 5

4 VARIANCE RESULTS

As described in more detail in Section 2, we estimate the signal standard deviation in prosecutor effects within sentencing strata — which are defined as a crime-type within an office and 5 year-block. To do this, we first split our sample in half, then compute each prosecutor’s average outcome vis-a-vis their sentencing strata in each split, and finally estimate the covariance in prosecutors’ outcomes across splits. This procedure isolates the component of the observed variation in prosecutors’ outcomes that replicates in independent samples. This is the component that can be attributed to prosecutor effects — the signals — rather than the draw of particular cases — the sampling error. We bootstrap this procedures to produce standard errors.

Variation in Incarceration Effects Figure 6 displays the point estimates (in the blue dots) and confidence intervals (the black horizontal lines) of the signal standard deviations of prosecutor incarceration effects within 6 months, 1 year, and 3 year windows. In the 6-months following disposition, for instance, a prosecutor one standard deviation above the mean in terms of her short-term incarceration effect tends to incarcerate defendants for an additional 7.8 days relative

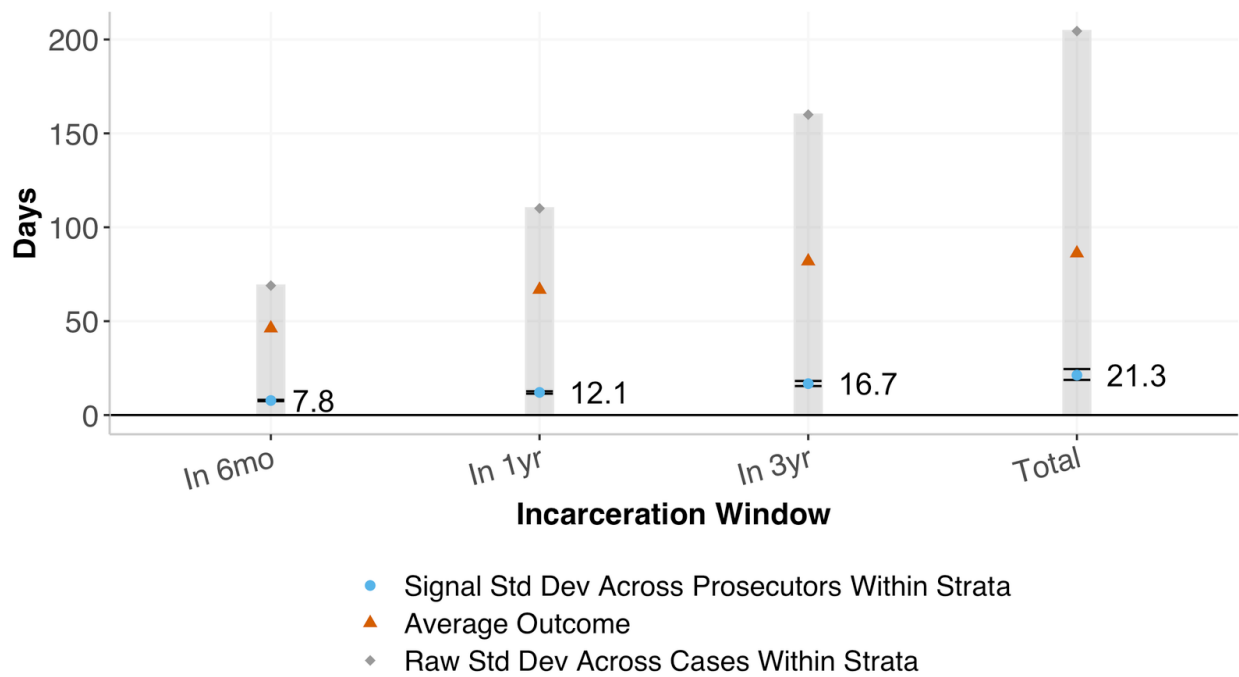


Figure 6: The blue dots represent the point estimates of the signal standard deviations of prosecutor incarceration effects within 6 months, 1 year, and 3 year windows. The signal standard deviation is the component of the observed variation in prosecutors' outcomes that replicates in independent samples. The black horizontal lines show the confidence intervals for the signal standard deviations, which are obtained by bootstrapping our procedure to estimate the signal standard deviations. To help contextualize the magnitude of this variation, the orange triangles depict the mean incarceration lengths within each window, and the grey bars depict the total, raw standard deviation in outcomes within a sentencing strata. We construct each of these statistics at the sentencing strata level before aggregating to produce a single statistic.

to other prosecutors in her sentencing strata. And a prosecutor one standard deviation above the mean in her incarceration effect within 1 year imposes sentence lengths that are 12.1 days higher than the average prosecutor in her sentencing strata.

To help contextualize the magnitude of variation across prosecutors, Figure 6 also depicts the mean incarceration length within each window in orange triangles and the raw standard deviation in outcomes within a sentencing strata in the grey bars. To compute the means and standard deviations, we construct these statistics at the sentencing strata level before aggregating to produce a single statistic. We can see that increase in the signal standard deviation in levels as the incarceration window increases roughly tracks the increase in the incarceration means as the incarceration window increases. Across the different windows, variation across prosecutors explains 10 to 11.5% of the total variation in sentencing outcomes within each sentencing strata – i.e. within an office, at a given time, and for a specific kind of crime. In the shorter time windows, prosecutor variation accounts for a slightly higher share of the total variation.

Figure 7 breaks down the heterogeneity in incarceration length attributable to prosecutors by demographic group.²⁸ The first panel depicts the aggregate variation across prosecutors. A prosecutor one standard deviation above the mean imposes sentences that are 17 days longer than the average prosecutor. This variation explains roughly 10% of the total variation in sentence length.

The second panel shows the variation in incarceration effects by race. We can see that a prosecutor one standard deviation above the mean in sentence length for black defendants imposes 15 additional days of incarceration. The heterogeneity in sentences for white defendants is similar. Since severity for black and white defendants only has a correlation of 0.1, the heterogeneity in the race gap is larger — a gap between black and white defendants that is 18.1 days higher than the average prosecutor’s gap. Since the average race gap in sentencing is 26 days, this variation seems especially pronounced (at 70% of the average). In future work, we will explore whether the tails of the race gap distribution are driving this large variation across prosecutors.

²⁸This analysis limits to the first three years following disposition to reduce the effects of outliers.

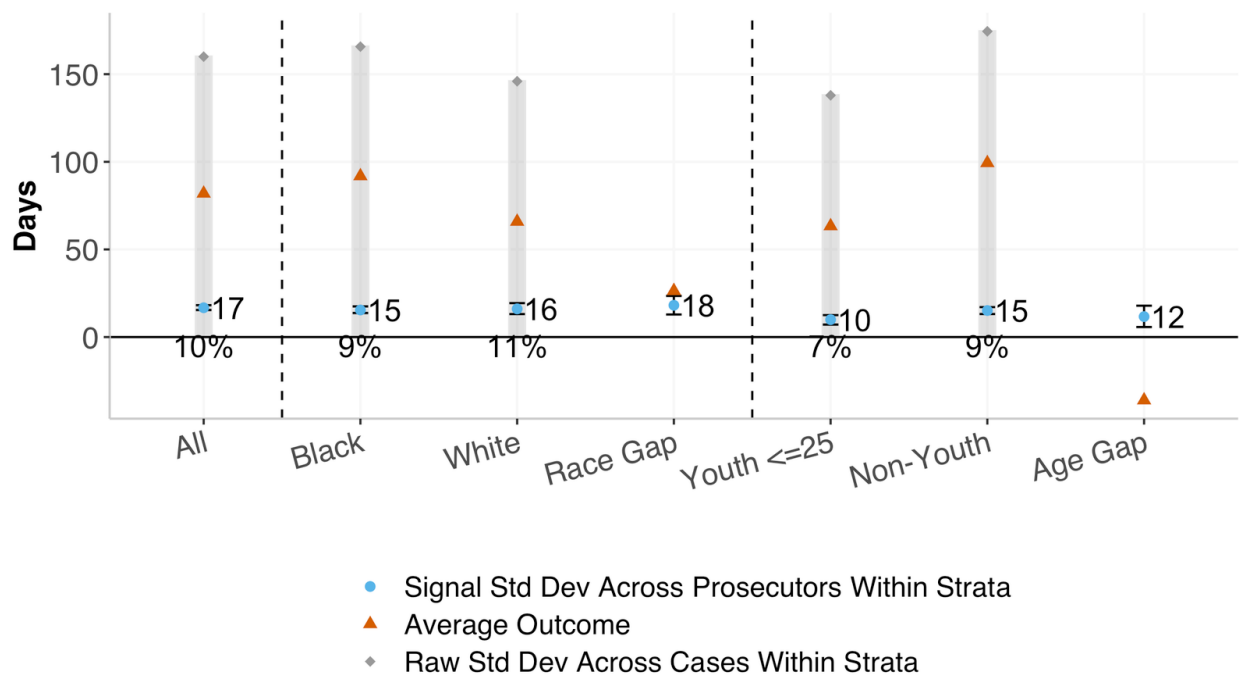


Figure 7: This figure breaks down prosecutor heterogeneity in incarceration length by demographic group. This analysis limits to the first three years following disposition to reduce the effects of outliers. The blue dots are the points estimates of the signal standard deviations within each demographic group, and the black bars are the bootstrapped standard errors. The red triangles show the means and the grey diamonds the total standard deviations for each outcome. The first panel depicts the aggregate variation across prosecutors. The second panel shows the variation in incarceration effects by race. Finally, the third panel shows heterogeneity in incarceration effects for defendants who are less than 25 or greater than 25 at the time of first charging.

The third panel shows heterogeneity in incarceration effects for defendants who are less than 25 or greater than 25 at the time of first charging. While the variation in incarceration effects for defendants older than 25 is similar in magnitude to the aggregate variation and variation by race, the point estimate for the variation in effects on young defendants — as well as the share of the total variation in sentencing for young defendants attributable to prosecutors — is markedly lower. Given the oft-cited desire to extend "second chances" to young defendants, this result is perhaps unsurprising.

Variation in Recidivism Effects and Skill Prosecutors also vary systematically in their recidivism effects. Figure 8 depicts the heterogeneity in prosecutor recidivism and "skill," where again the blue dots represent the heterogeneity across prosecutors, the orange triangles represent the aggregate means, and the grey diamonds represent the total standard deviations in each outcome. In the first bar, we see that a prosecutor with a re-arrest rate in her cases²⁹ that is one standard deviation above the mean has a re-arrest rate that is 5.8pp higher than the average prosecutor, or 10.9% higher than average. The share of the total variation attributable to prosecutors is 12%, which is slightly higher than the share for incarceration outcomes. Given that prosecutors have more direct control over incarceration than recidivism outcomes, this result is perhaps unexpected.

The variation in prosecutors' re-conviction rates is slightly lower than for re-arrests but represents a higher fraction of the mean re-conviction rate and a comparable fraction of the total variation. The re-conviction effect includes probation violations that do not result in a conviction (e.g. when a defendant violates the terms of his probation) while the re-arrest effect does not.

In the fourth bar, we can see that the variation in rearrest for violent crimes is about 2pp, which is substantially lower than for all crimes. However, the level of re-arrest for violent crimes is also substantially lower — about 4% — and therefore, a prosecutor one standard deviation above the mean has a 50% higher rearrest rate for violent crime than the average prosecutor.

The third bar shows the variation in prosecutor skill, where skill is estimated by netting out from

²⁹Arrests are measured within three years of the case disposition date.



Figure 8: This figure captures the heterogeneity in prosecutor effects on re-arrest and heterogeneity in prosecutor “skill.” The blue dots represent the heterogeneity across prosecutors, the orange triangles represent the aggregate means, and the grey diamonds represent the total standard deviations in each outcome. In the first bar, for instance, we see that a prosecutor with a re-arrest rate in her cases (within three years of case disposition) that is one standard deviation above the mean has a re-arrest rate that is about 6pp higher than the average prosecutor. The second and fourth bars capture heterogeneity in re-conviction rates and rates of violent re-arrest. The third bar captures heterogeneity in prosecutor skill, where skill is estimated by subtracting from a prosecutor’s re-conviction effect, the expected re-conviction effect given a prosecutor’s incarceration effect and the aggregate relationship between incarceration and re-conviction

a prosecutor's recidivism effect their expected re-conviction effect (given their incarceration effect and the aggregate relationship between incarceration and re-conviction). Put differently, this skill measure attempts to eliminate the the average mechanical, incapacitation bump that prosecutors receive from elevated rates of incarceration. The variation in skill across prosecutors is similar in magnitude to the variation in unconditional re-arrest and re-conviction rates: a prosecutor one standard deviation above the mean in skill can achieve a 4.6pp 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 suggest that there exists positive variation in prosecutor skill even if skill were homogeneous. 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.³⁰³¹

To provide intuition for how these heterogeneity results might translate into changes in case outcomes in practice, we consider the thought experiment of selectively retaining prosecutors on the basis of their observed incarceration and reconviction effects. In finite data, such an exercise would inevitably lead to mistakes, as some prosecutors with a lucky draw of cases would appear

³⁰A preliminary analysis suggests that this frontier may be well approximated by a linear function.

³¹In 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.

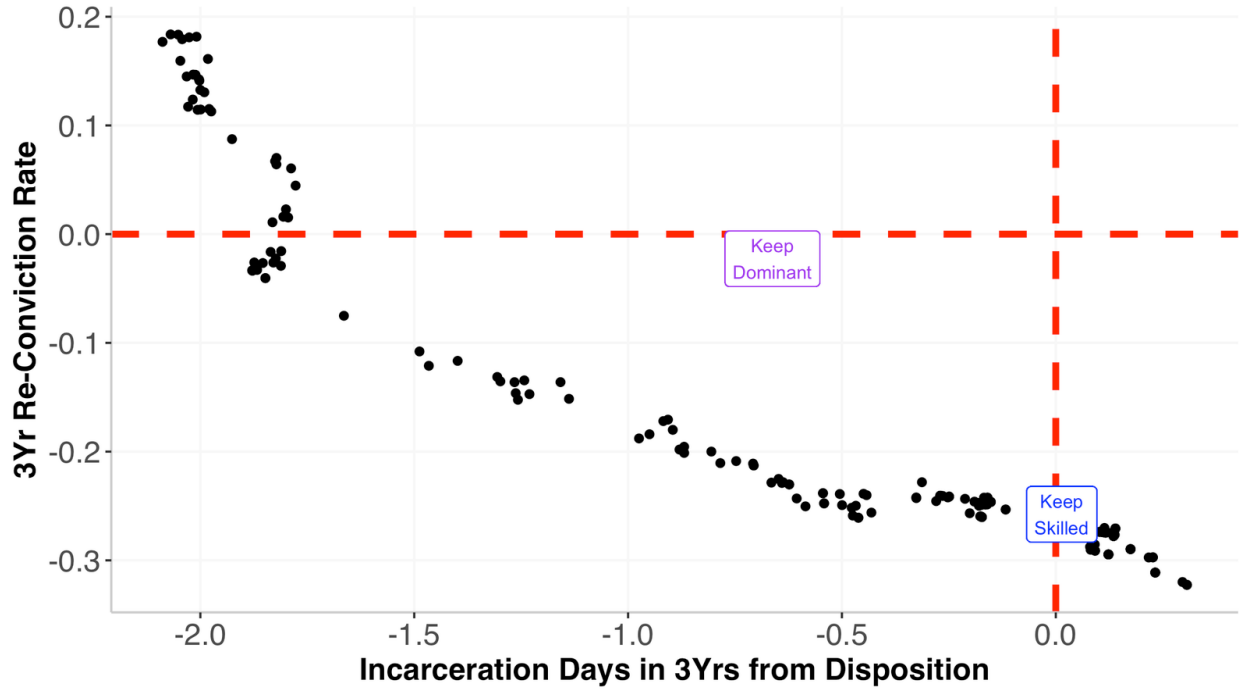


Figure 9

better than their true effects. This "winner's curse" bias in turn would lead to poorer results out of sample. To eliminate this bias, we evaluate prosecutors on one split of the data and use the resulting outcomes in the other split of the data for this policy exercise. Figure 9 traces out the resulting frontier of incarceration and re-conviction outcomes as one varies the weights on these two outcomes and retains prosecutors according to their weighted combination of incarceration and re-conviction effects.

Figure 9 also highlights the results of keeping (a) "skilled" prosecutors — those who have lower re-conviction rates than would be expected given their incarceration rates and (b) "dominant" prosecutors — those who have both lower re-conviction and lower incarceration rates than other prosecutors in their sentencing strata. Keeping the skilled prosecutors reduces reconviction rates without changing incarceration rates.

5 COVARIANCE RESULTS

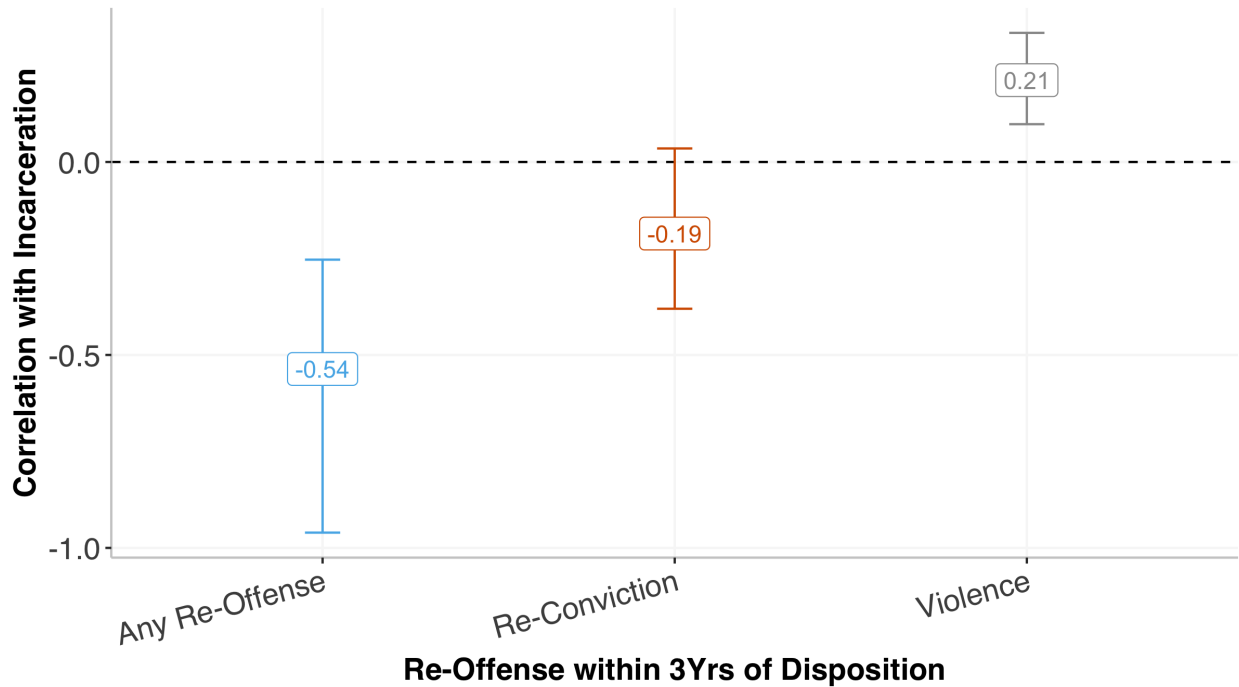


Figure 10: This figure shows the correlation between prosecutor recidivism effects and prosecutor's incarceration effects. Incarceration is truncated at three years to reduce the influence of outliers. As in our variance estimates, we similarly leverage a split sample approach to eliminate the contribution of sampling error to our estimates.

The relationship between Incarceration and Recidivism Effects As expected, given the incapacitation effect and the aggregate frontier from figure 9, figure 10 confirms that the correlation between prosecutors' incarceration and re-convictions effects is negative (-.19). Despite the predictable impact of the incapacitation effect, however, this means that only 29% (4%) of the systematic variation in prosecutors' re-arrest (re-conviction) 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 remaining 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, allowing some prosecutors to beat the aggregate relationship between incarceration and re-offense.

Given this result, it is less surprising that variation in prosecutor skill (which adjusts for the prosecutor's incarceration effect) is similar in magnitude to the estimated variation in prosecutor effects on re-offense. As shown in Figure 8, a one standard deviation movement in the distribution of

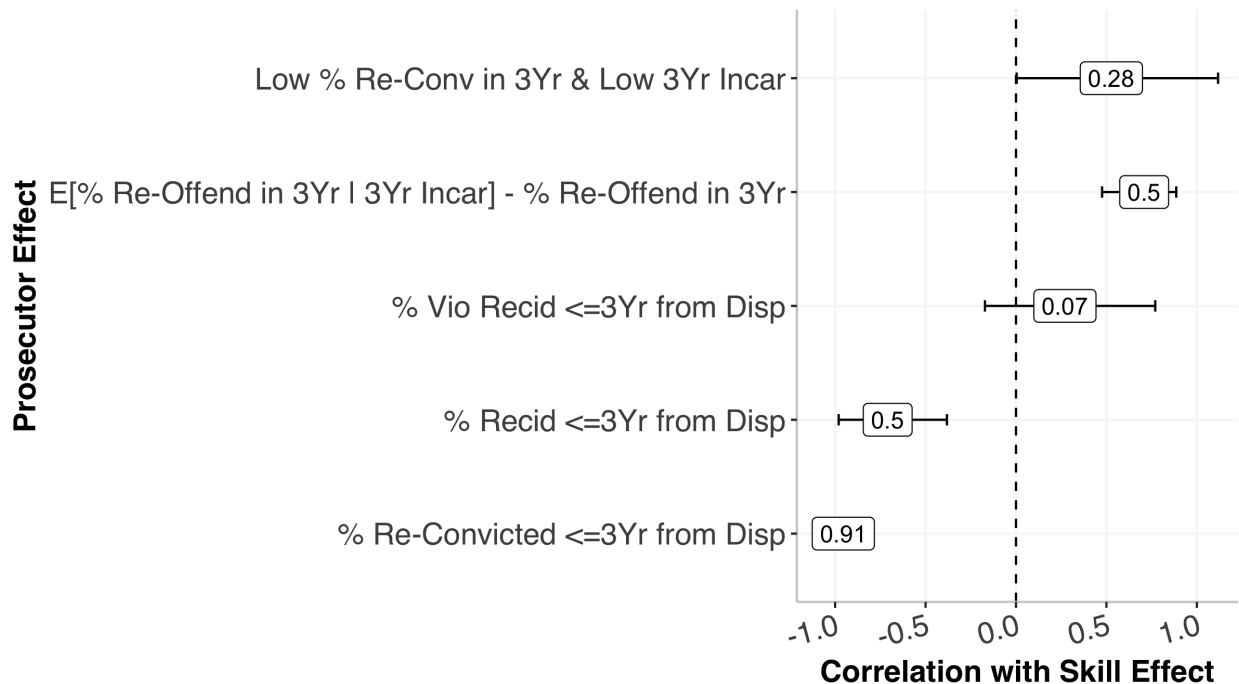


Figure 11: This figure shows the relationship between several measures of prosecutor skill and prosecutor's recidivism effects. The top measure is an indicator that the prosecutor is below average in her incarceration and recidivism effect. The second measure is prosecutor skill with respect to re-arrest as opposed to re-conviction. The final three bars show the correlation of prosecutor skill with prosecutors' recidivism effects (that do not account for the predictable impact of incarceration on recidivism).

re-conviction versus skill induces a 4.65pp lower rates of re-conviction unconditionally versus a 4.57pp lower rates of re-conviction than would be expected given her incarceration effect.)

Even more surprisingly, Figure 10 reveals that prosecutors who impose longer sentences tend to have *higher* rates of violent recidivism, suggesting that those prosecutors who reduce violent re-offense the most tend to be more lenient in aggregate.

Comparing different measures of skill. As shown in figure 11, skilled prosecutors are also more likely to be "dominant" — that is, to have below average incarceration and below average re-conviction rates. The correlation between these two measures of prosecutor skill is 0.61 — and so 38% of the variation in dominance can be explained by variation in our definition of skill.

Prosecutors with high estimated skill using the metric of re-conviction also tend to be highly skilled when assessed on the metric of re-offense of any kind, including probation violations that

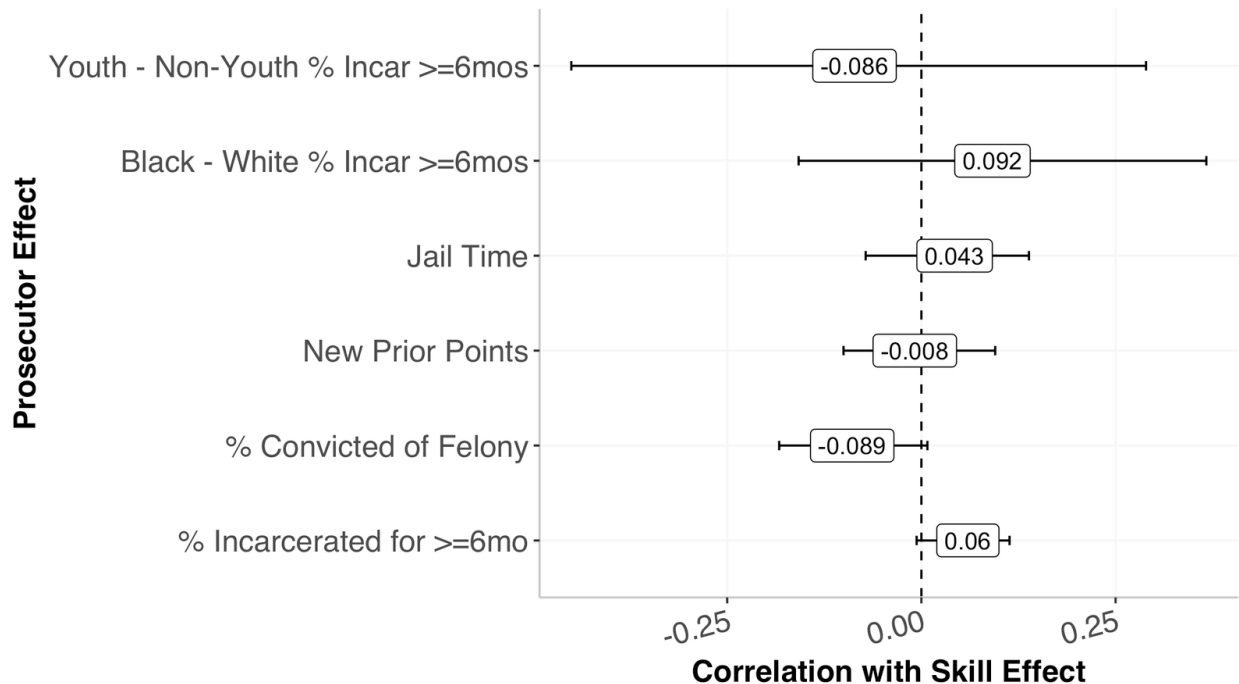


Figure 12: This figure presents the relationship between skill and other prosecutor effects. The top two bars show the relationship with prosecutors' effects on the race gap and age gap. The third bar shows the relationship with a prosecutor's tendency to impose jail time, which often occurs prior to the case's resolution. The fourth bar measures the relationship between prosecutor skill and the prosecutor's effect on the severity of the final conviction charge — which determines the number of additional "prior points" on a defendant's record were he to be charged and sentenced in the future. The fifth bar measures the relationship with a prosecutor's tendency to convict defendants of felony charges. Finally, the bottom bar shows the relationship with a prosecutor's tendency to impose incarceration sentences of at least six months.

do not lead to significant prison sentences. However, the correlation in these metrics of skill is only 0.64, and 59% of the variation in re-offense (relative to what would be expected given a prosecutor's incarceration effect) cannot be explained by our skill measure. This suggests that certain prosecutors may weight the risk of a probation violation more than other prosecutors when determining whom to incarcerate. We also find that our measure of skill is negatively related (with a correlation of $-.2$) to a prosecutor's effect on violent re-offense in a prosecutor's caseload.

Unpacking Skill: the relationship between skill and other prosecutor effects

Figure 12 presents the relationship between skill and other prosecutor effects. The top two bars reveal that there is no statistically significant pattern between skill and incarceration gaps for race or age.

The bottom three bars suggest that skilled prosecutors tend to press for less severe punishments. Starting with the very bottom bar, we see that skilled prosecutors are more likely to impose incarceration sentences of at least six months. Since our measure of skill removes the component of recidivism that is expected given the prosecutor's incarceration rate — and so effectively controls for a prosecutor's incarceration length — this covariance implies that skilled prosecutors must impose shorter sentences. In order to achieve incarceration rates equal to those prosecutors with lower incarceration rates, prosecutors with higher incarceration rates must have a compressed distribution of incarceration sentences.

This result, coupled with two bars above, suggests that skilled prosecutors tend to impose more frequent but more lenient punishments. Skilled prosecutors are less likely to convict defendants of a felony and more likely to drop down to a misdemeanor. They also tend to convict defendants of less severe felonies (ones that are associated with fewer prior points). These relationships are consistent with shorter, lighter sentences being effective for reducing recidivism. By contrast, long sentences and felony convictions may be counter-productive in reducing recidivism. This could be due to diminishing value of the incapacitation effect as sentences increase (perhaps due to defendants "aging out" of crime). This may also be due to the more severe collateral consequences of incarceration, which tend to kick in more for felony convictions than misdemeanors — and perhaps more for longer prison sentences. For instance, a convicted felon or a defendant who serves a significant prison sentence may face steeper hurdles to employment, thereby increasing the likelihood of re-conviction. In future work, we hope to explore this relationship further.

6 APPENDIX: IDENTIFYING SELECTION

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:

$$Y_{ipgct} = \mu_p + \mu_{gct} + X_i + \epsilon_{ipgct}. \quad (5)$$

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 5 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 ap-

proach in equation 5 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.³² In this way, the cases that link different crime-types via a shared prosecutor are given outsized influence in equation 5. 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.^{33 34}

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 be-

³²Further, 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.

³³In 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.

³⁴Our 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.

coming 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 5, but restrict to offices that do not have units with offense type specialization.

7 APPENDIX: ANALYTICAL CORRECTIONS

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

$$\mathbb{E}[\bar{Y}_{pgct} - \bar{Y}_{gct}] = \mu_{pgct} - \bar{\mu}_{gct} + \underbrace{\mathbb{E}[\bar{\epsilon}_{pgct} - \bar{\epsilon}_{gct}]}_0.$$

However, a collection of unbiased estimates of individual prosecutor effects can still generate biased estimates of the heterogeneity in their effects.

The raw variation in prosecutors' effects will include one component, the signal variation, that reflects the heterogeneity in prosecutors' causal effects and component, the noise variation, that reflects each prosecutor's particular random draw of cases:

$$\mathbb{E}[\sum_{pgct} (\bar{Y}_{pgct} - \bar{Y}_{gct})^2] = \sum_{pgct} (\mu_{pgct} - \bar{\mu}_{gct})^2 + \underbrace{\mathbb{E}[\sum_{pgct} (\bar{\epsilon}_{pgct} - \bar{\epsilon}_{gct})^2]}_{\neq 0}. \quad (6)$$

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 6. Subtracting this from the raw variation recovers an unbiased assessment of the true variation in prosecutors' effects. When we instead estimate the covariance in prosecutor effects — in e.g., incarceration and re-offense — we can correct the raw covariance with the comovement in estimation errors across the two estimating equation. This accounts for the comovement that one would expect due to sampling error alone.