NBER WORKING PAPER SERIES

RACIAL DIVISIONS AND CRIMINAL JUSTICE:
EVIDENCE FROM SOUTHERN STATE COURTS

Benjamin Feigenberg
Conrad Miller

Working Paper 24726
http://www.nber.org/papers/w24726

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2018

Eric Hernandez and Colin Ash provided excellent research assistance. We thank Amanda Agan, Anil Jain, Jonathan Leonard, Aurelie Ouss, Evan Rose, Yotam Shem-Tov, and seminar participants at Pomona College, UC Berkeley, UCLA, University of Illinois at Chicago, University of Chicago Crime Lab, and University of Chicago Harris School for comments. We thank the Center for Science and Law for providing access to Alabama court data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Racial Divisions and Criminal Justice: Evidence from Southern State Courts
Benjamin Feigenberg and Conrad Miller
NBER Working Paper No. 24726
June 2018
JEL No. J15,K14,K41,K42

ABSTRACT

The US criminal justice system is exceptionally punitive. We test whether racial heterogeneity is one cause, exploiting cross-jurisdiction variation in criminal justice practices in four Southern states. We estimate the causal effect of jurisdiction on initial charge outcome, validating our estimates using a quasi-experimental research design based on defendants that are charged in multiple jurisdictions. Consistent with a simple model of ingroup bias in electorate preferences, the relationship between local punitiveness and the black share of defendants follows an inverted U-shape. Heterogeneous jurisdictions are more punitive for both black and white defendants. By contrast, punishment norms are unrelated to local crime rates. Simulation results suggest that adopting the punishment norms of homogeneous jurisdictions would decrease the share of charges leading to an incarceration sentence and the black-white gap in this share by 16-19%.

Benjamin Feigenberg
Department of Economics
University of Illinois at Chicago
601 South Morgan Street
706 University Hall
Chicago, IL 60607
bfeigenb@uic.edu

Conrad Miller
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720
and NBER
ccmiller@berkeley.edu

1 Introduction

The United States incarcerates residents at a higher rate than any other country in the world. While less than five percent of the world's population resides in the US, nearly twenty five percent of the world's prison population is held in US facilities (Walmsley, 2016). Though differences in violent crime rates can in part explain this pattern, the US is also exceptionally punitive (Pfaff, 2014). Some observers have argued that race plays a key role in driving American criminal justice policy (Alexander, 2010). There is prima facie evidence: US blacks are incarcerated at six times the rate of whites and face longer sentences for similar crimes (Carson, 2014; Rehavi and Starr, 2014). Race may play a broader role, even influencing the incarceration rate for US whites, which itself would rank near the top among developed nations (Gottschalk, 2015). Just as racial heterogeneity predicts lower support for redistribution and public goods (Alesina et al., 1999), it may increase support for harsher punishment if, for example, voters prefer to punish outgroup members more severely. In this paper, we ask whether racial heterogeneity can in part explain US exceptionalism in criminal justice.

Empirical research on the role of race in criminal justice policy is complicated by the difficulty of separating the relative importance of policy versus underlying criminal conduct in generating cross-country variation in incarceration rates. Harmonized micro data covering the US and a significant number of other countries do not exist, and differences in the definitions of crimes across countries would make harmonization extremely difficult. Instead, we study the relationship between racial divisions and criminal justice policy by investigating cross-jurisdiction variation in punishment *within* US states. In doing so, we take advantage of harmonized data and fixed criminal codes within states and exploit the substantial within-state variation in how criminal law is enforced.

While much statutory criminal justice policy is driven by state-level legislation, localities have significant discretion in how they enforce those laws, and that discretion is tied to electorate preferences. Prosecutors and judges are often locally elected and influence outcomes at each stage of the criminal justice process: prosecutors decide what charges to file; prosecutors negotiate plea bargains; judges make sentencing decisions after conviction. The electorate may also affect adjudication outcomes by serving as jurors or influencing spending on indigent defense. A 2016 New York Times Upshot article illustrates the role of local politics in driving local punitiveness with a quote from the elected prosecutor in Dearborn County, Indiana: "I am proud of the fact that we send more people to jail than other counties...My constituents are the people who decide whether I keep doing my job. The governor can't make me. The legislature can't make me" (Keller and Pearce, 2016).

In this paper, we evaluate the role that racial heterogeneity plays in determining criminal justice outcomes. We first estimate local *punishment norms*, the causal effect of jurisdiction on the outcome of an initial criminal charge, using data from four Southern states. We then link variation in punishment norms to local racial heterogeneity in the population. Consistent with a simple model of ingroup bias in electorate preferences, we find that the relationship between local punitiveness and black defendant share follows an inverted U-shape: jurisdictions with the largest white and black majorities are relatively lenient while intermediate, heterogeneous jurisdictions are more punitive.

To measure punishment norms, we use rich criminal justice administrative data that track criminal cases from initial charging through sentencing, including dropped charges. To credibly isolate across-jurisdiction variation in sentencing outcomes that is explained by local norms, we adopt methods from the teacher value-added (Chetty et al., 2014) and worker-firm wage decomposition (Abowd et al., 1999; Card et al., 2013, 2016) literatures. We employ a quasi-experimental research design that exploits variation in outcomes for defendants arrested in multiple jurisdictions.[1] We validate our punishment norm estimates by showing that they accurately predict the within-defendant *changes* in charge outcomes coinciding with changes in jurisdiction. Throughout the analysis, our benchmark specifications focus on the share of charges that lead to incarceration sentences (the *confinement rate*) as the relevant measure of punitiveness, though we present supplementary estimates to confirm that our findings are robust to case-level rather than charge-level specifications.

Our data come from Alabama, North Carolina, Texas, and Virginia, which account for about 20% of all prisoners held under state jurisdiction in the US. We focus on the South because there is substantial variation in racial composition across Southern counties. Across counties in the states we study, the black share of the population ranges from 0% to 85%. In all four states, district attorneys are locally elected; in all but Virginia, judges are locally elected. The data reveal substantial within-state heterogeneity in jail and prison admissions: across states, the coefficient of variation for jurisdiction-level admissions per capita ranges from 33% to 58% in our sample.

The variation we measure in admissions per capita is matched by substantial heterogeneity in punishment norms: the coefficient of variation for punishment norms ranges from 25% to 37% across the four states in our sample. A defendant charged in a jurisdiction in the top quartile by punitiveness is 2-3 times more likely to be incarcerated for a given charge than a defendant charged in a jurisdiction in the bottom quartile. We show that punishment norms explain 67-79% of the within-state, across-jurisdiction variation in confinement rates.

Interestingly, punishment norm estimates constructed separately by race are highly correlated. Jurisdictions that are more punitive for black defendants are also more punitive for white defendants. We also find substantial variation across jurisdictions in the rate at which prosecutors pursue charges. This rate is positively correlated with punishment norms, suggesting that prosecutors are a key driver of those norms.

[1] In the corresponding worker-firm wage decomposition and teacher value-added literatures, researchers exploit workers that move from one firm to another and teachers that move from one school to another.

We next document the relationship between local punishment norms and racial heterogeneity. We motivate our analysis with a simple model of ingroup bias where voters prefer more severe punishment when offenders are more likely to belong to a different racial group. Prior work documents that common group membership is associated with declines in envy and punishment for misbehavior as well as increases in charitable concerns and rewards for good behavior (Chen and Li, 2009). This mechanism suggests that the relationship between local punishment norms and the black share of defendants (or general population) will follow an inverted U-shape; while white voters prefer more punitive policy as the black share of defendants increases, for jurisdictions with larger black populations, the pivotal voter is more likely to be black. This prediction is borne out in the data.

Surprisingly, lagged growth in violent crime, which has been previously identified as an important driver of cross-state variation in incarceration rates (Western, 2006), is uncorrelated with severity in our sample. Following the cross-state analysis presented in Western (2006), we introduce additional covariates in alternative specifications, including average income, income inequality, Republican vote share, the fraction of prime-aged males in the population, and population density. Among these covariates, population density and Republican vote share both consistently predict higher confinement rates.

We conclude by simulating outcomes under a counterfactual in which all jurisdictions adopt the severity level imposed by those jurisdictions with racially homogeneous populations. Under this counterfactual, we show that overall confinement rates and racial confinement rate gaps fall by 16-19% once we account for both the static effect of lower punitiveness on confinement outcomes and the dynamic effect of lower punitiveness on defendants' criminal histories.

In emphasizing the importance of racial divisions as a key driver of electoral preferences and local punitiveness, we build on a large literature that highlights the racialized nature of crime in the US (Muhammad, 2010) and the role of 'racial threat' in explaining policy and punishment preferences (Key, 1949; Glaser, 1994; Enos, 2015; Unnever and Cullen, 2007). The racial threat literature studies how the presence of racial and ethnic minority populations affects white voting behavior and policy preferences. While findings in this literature are generally inconsistent, the most recent and compelling evidence suggests that a larger minority population increases white voter turnout and support for conservative policies and candidates (Enos, 2015).

A related literature uses survey data to link white racial attitudes to criminal justice preferences. Whites who express more racial resentment are more likely to support capital punishment and other harsh crime-control policies (Unnever and Cullen, 2010). Respondents who are primed to consider the prison population as 'more black' express more concern about crime and greater acceptance of punitive policies (Hetey and Eberhardt, 2014). There is evidence that public support for 'tough on crime' policies tracks national incarceration rates over time (Enns, 2014). While we cannot measure local preferences directly, we measure local policy in the form of punishment norms.

Motivated by the racial threat hypothesis, several papers test for a relationship between state racial composition and imprisonment rates, with mixed results. Most relevant to our work, Keen and Jacobs (2009) finds an inverted U-shaped relationship between black population share and racial *disparities* in state prison admissions per capita. In contrast with this past research, we focus on county-level criminal justice and we employ a mover-based identification strategy to credibly isolate the causal effect of charge location on sentencing. We find an inverted U-shaped relationship between county black population share and punishment norms that applies to all defendants. By contrast, we find no evidence that local racial composition is correlated with local racial disparities in punishment severity.

In using residual confinement rate differences (conditional on defendant and case covariates) to estimate punishment norms, we build on Rehavi and Starr (2014), which identifies substantial racial disparities in federal sentencing outcomes.[2] While the importance of local discretion has received increased popular attention in recent years, to the best of our knowledge, ours is the first work to comprehensively characterize the degree to which local criminal justice system actors drive cross-jurisdiction variation in defendant outcomes. Understanding those factors that influence cross-jurisdiction state court punishment norms is highly policy-relevant given that state prison and jail admissions account for over 90% of all admissions in the U.S.

Our paper builds on a literature that studies the role of electoral pressure on the composition and behavior of judges and, to a lesser extent, prosecutors. Broadly, this literature finds that judge and prosecutor behavior is tied to local electorate preferences. Voters elect like-minded public officials, or public officials are responsive to electorate preferences. Huber and Gordon (2004) and Berdejó and Yuchtman (2013) find that judges in Pennsylvania and Washington sentence serious crimes more severely when they come up for reelection. The authors argue that this pattern of judge behavior is driven by re-election incentives and the preferences of the electorate. Lim (2013) finds that judges in counties using partisan judicial elections exhibit different sentencing patterns from judges in counties using referendum judicial elections, and attributes these differential patterns to differences in electoral pressure under the two systems. Lim et al. (2015) find that newspaper coverage increases sentence length by nonpartisan elected judges for violent crime, and argue this relationship is mediated through electorate preferences. There is also evidence that prosecutors respond to constituent preferences (Dyke, 2007; Nelson, 2014). In our model, the predicted relationship between local punishment norms and racial

[2] In that work, the authors employ federal criminal records data that allow them to eliminate sample selection concerns that arose in previous analyses which conditioned on defendant conviction. The small number of prior studies that analyzed the emergence of race-based disparities in intermediate outcomes, such as plea bargaining, did not condition on key covariates such as defendant criminal history and specific offense. As a result, estimated disparities may be race-based or may be driven by differences in unobserved characteristics of criminal defendants. See, for instance, Miethe (1987). It is important to note that the extent of selection bias may be more limited in federal criminal cases than in state cases. Fischman and Schanzenbach (2012), for instance, conditions on conviction but argues that associated selection bias is limited because acquittals account for only one percent of the federal criminal cases that they analyze.

composition that we document is mediated through electorate preferences.

Our work also contributes to a public finance literature that studies the association between local racial composition and policy preferences. In Alesina et al. (1999), the authors provide evidence that public goods spending is inversely related to ethnic fragmentation in US cities and argue that this finding is driven by cross-group policy preference heterogeneity. Luttmer (2001) shows that self-reported support for welfare spending is increasing in the share of local recipients from the respondent's own racial group. The link between ethnic fragmentation and support for redistribution is also buttressed by Dahlberg et al. (2012), which shows that plausibly exogenous increases in immigration to Swedish municipalities are associated with decreases in support for redistribution. We argue that the inverted U-shape relationship between black defendant share and severity of incarceration policy in our data can be explained by the same racial group loyalty that drives the positive association between racial homogeneity and support for redistribution.

The remainder of the paper is structured as follows. Section 2 describes the data used for the analysis. Section 3 discusses our approach to characterizing cross-jurisdiction differences in punishment norms, including our mover-based identification strategy, and provides estimates. Section 4 presents a model of racial group loyalty to highlight the role that racial divisions may play in explaining this variation and empirically tests the predictions of the model. In Section 5 we simulate counterfactual confinement rates. Section 6 concludes.

2 Data

We use administrative criminal justice data from four states: Alabama, North Carolina, Texas, and Virginia.[3] The data source and years of data we analyze for each state are presented in Table 1. We summarize the content of the data here and discuss data construction and state-specific institutional context in greater detail in the Data Appendix.

Though the data from each state differ in their exact content, they all track state felony and misdemeanor criminal cases from initial charging through sentencing, and share important data elements. Critically, data for all states include charges that are ultimately dropped. Data from all states include information on each arrest charge, including the arrest offense, date of arrest, the court where the case is assigned, final court charge, case disposition, and, if the case results in conviction, the final sentence. Defendant information includes date of birth (except Virginia, which does not include year of birth), gender and race. Data from North Carolina and Texas also identify Hispanic defendants.

For all states, the data include property, violent, and drug offenses. We refer to these offenses categories as 'Core' offenses. The data also include 'crimes against society', including driving

[3] We have also analyzed data from Arkansas and Maryland. However, we omit data from these states due to data quality issues. Including data from these states does not substantively affect any of the results reported in this paper.

Table 1: Data by State

State	Source	Year
Alabama	Alabama Administrative Office of the Courts	2000-2010
North Carolina	North Carolina Administrative Office of the Courts	2007-2014
Texas	Texas Department of Public Safety	2000-2010
Virginia	Virginia's Office of the Executive Secretary	2006-2014

Notes: Data sources are discussed in more detail in the Data Appendix.

while intoxicated (DWI), writing bad checks, and trespassing. For all states, we drop non-DWI traffic offenses. We also exclude charges in which the final listed disposition is an intermediate outcome, such as a transfer between district and circuit courts or across jurisdictions. Lastly, we exclude probation and parole violations. While we include all remaining charges in the baseline analysis, we also explore limiting charges to Core offenses as a robustness check.

For all states, we drop cases for defendants aged below 16, which are likely to be adjudicated within the juvenile justice system. We also exclude offenses with fewer than 100 occurrences in the data and offenses with zero instances that result in confinement.[4] These offenses are rare–the procedure removes many specific offense codes from the data, but only around 1% of charges. This leaves us with about 400-600 unique offenses in each state.

In Alabama and Virginia, we restrict to black and white defendants. In Alabama, American Indian-, Asian-, and Hispanic-coded defendants account for less than 0.25% of charges. In Virginia, the same categories amount for about 2% of charges. In North Carolina and Texas, we restrict to black, white, and Latino defendants. American Indian- and Asian-coded defendants account for less than 2% and 1% of charges in these states, respectively. In all states, we drop defendants with missing race codes. These account for about 1% or less of charges in all states.

In defining jurisdictions within states, we consider the county the relevant unit of analysis. All results are similar if we instead use prosecutorial or judicial districts, which sometimes include multiple contiguous counties.

2.1 Confinement Sentence as a Benchmark Outcome Measure

There are several potential outcomes to use for measuring punishment norms. A criminal charge can be pursued or dropped by the prosecution. Pursued charges can result in conviction, acquittal, deferred judgment, or some other outcome. Conviction can lead to probation or confinement sentences of varying lengths, or an alternative sentence.

For our measure of severity, we focus on whether a given charge results in a confinement sentence. That is, we focus on whether a charge results in an unconditional jail or prison

[4] When we analyze court outcomes at the case-level rather than the charge-level as described below, we include these offenses when constructing controls if they are not the primary charge in the case.

sentence. We focus on this measure for two reasons. First, there is evidence that the extensive margin of incarceration is critical for recidivism and labor market outcomes (Mueller-Smith, 2015). Second, our data generally do not include information on the mapping between sentence length and realized sentence, which may vary across jurisdictions in ways we cannot measure. The same nominal sentence in two counties may lead to different realized sentences, for example, due to sentence reductions or enhancements, or parole board decisions.

2.2 Descriptive Statistics

We tabulate descriptive statistics for charge data from each state in Table 2. We include information on defendant demographics, charge characteristics, and charge outcomes. 'Long Sentence' is defined as a confinement sentence of at least 360 days. The Alabama data do not indicate whether a charge is a felony or misdemeanor, and we only include Latino defendants in North Carolina and Texas.[5]

The number of charges in our data ranges from 1.9 million in Alabama to 5.9 million in Texas. The number of charges per defendant ranges from 2.3 to 2.6. Across states, 71.1-78.7% of charges are filed against male defendants. Defendants are disproportionately black; while the black share of the population ranges from 11.8% in Texas to 26.1% in Alabama, the black share of defendants ranges from 24.4% in Texas to 41.8% in North Carolina. In both Texas and North Carolina, the Latino share of defendants is *lower* than the Latino share of the population. However, there is evidence that law enforcement may underreport Latino status (Collister, 2015).

In the states where felony status is recorded, 27.0-35.9% of charges are felonies. The distribution of offense types varies across states, though in each state a plurality of charges are for 'Other' offenses.

Note that charge outcomes vary significantly across states. This is due in part to variation in severity across states, but may also be due to differences in charging behavior across states. Across states, the same crime may result in a different set of arrests, which may in turn result in a different set of recorded charges.[6] Throughout the analysis we focus on comparing jurisdictions within states.

We compare jail and prison admissions across jurisdictions within states in Table 3. We use three measures: jail and prison admissions per 100,000 residents, jail and prison admissions per case, and the share of charges that lead to a jail or prison sentence. Throughout, we refer to the last measure as the *confinement rate*. While the first measure incorporates variation in number of cases and charges per capita across jurisdictions, the second and third measures come closer to capturing how a given case or charge is treated differently across jurisdictions.

There is substantial variation in all three measures. For admissions per 100,000 residents, the

[5] See the Data Appendix for more details.

[6] Of course, charging behavior may across vary across jurisdictions within states, an issue we explore in more detail below.

Table 2: Descriptive Statistics

	Alabama	North Carolina	Texas	Virginia
Male	71.1	74.1	78.7	73.5
Black	37.2	41.8	24.4	40.7
Latino		4.4	30.4	
Age	32.9	31.8	31.0	
	(10.9)	(12.0)	(10.8)	
Felony		27.0	31.2	35.9
Property	17.0	31.5	22.1	30.3
Violent	10.0	14.5	12.2	9.9
Drug	17.3	16.9	21.8	13.4
Other	55.6	37.0	43.9	46.4
Dropped	40.4	61.5	22.3	40.9
Convicted	57.5	36.1	55.4	55.2
Probation	27.8	16.9	30.9	9.9
Confinement	27.1	9.0	40.2	16.1
Long Sentence	31.2	19.3	21.4	28.2
N Defendants	727,419	1,839,677	2,588,167	1,210,989
N Charges	1,854,208	4,770,146	5,875,045	2,897,055
N Cases	1,221,317	3,982,623	4,930,220	1,993,509
Charges per Defendant	2.5	2.6	2.3	2.4
	(4.3)	(4.1)	(2.4)	(4.2)
Cases per Defendant	1.7	2.2	1.9	1.6
	(2.0)	(2.7)	(1.7)	(1.9)

Notes: Missing values reflect outcomes unavailable for particular states. ‘Other’ offenses include crimes against society and offenses we are unable to classify due to miscoding. ‘Long Sentence’ is defined as a confinement sentence of at least 360 days, and is calculated *conditional* on any confinement.

Table 3: Jail and Prison Admissions Across Jurisdictions

	Alabama	North Carolina	Texas	Virginia
Admissions Per 100,000:				
Mean	609	546	789	696
SD	(358)	(179)	(331)	(374)
Admissions Per Case:				
Mean	0.249	0.107	0.405	0.201
SD	(0.100)	(0.031)	(0.153)	(0.058)
Confinement Sentence Per Charge:				
Mean	0.272	0.093	0.408	0.165
SD	(0.115)	(0.027)	(0.147)	(0.049)
N Jurisdictions	67	100	253	117

Notes: Statistics weighted by county population. 'Admissions per 100,000' is the total number of cases resulting in a jail or prison sentence in a county and year divided by county population in that year, averaged across years, and multiplied by 100,000. 'Admissions per Case' is the rate that cases result in a jail or prison sentence. 'Confinement Sentence Per Charge' is the rate that *charges* result in a jail or prison sentence.

coefficient of variation varies from 33% in North Carolina to 59% in Alabama. For admissions per case, the coefficient of variation varies from 29% in Virginia to 40.2% in Alabama.

3 Estimating Punishment Norms

A key objective of this paper is to measure and compare the severity of criminal punishment across jurisdictions. We posit that jurisdictions have *punishment norms*–they vary systematically in how they punish equivalent initial charges. Formally, we posit that charge outcomes take the following general form:

$$Y_{ict} = f(X_{ict}, j(i,c,t), \epsilon_{ict}) \tag{1}$$

where i indexes individuals, c indexes initial charge, and t indexes year. Y_{ict} is the outcome of the charge; we discuss the choice of our benchmark outcome measure in Section 2. The charge outcome is a function of X_{ict}, a vector that includes all relevant characteristics of the charge, including: the identity of the defendant, the specific charges filed, the date, and the quality of the evidence. The outcome of the charge is also a function of the jurisdiction in which the defendant is charged, $j(i,c,t)$. The relationship between the charge outcome and charge jurisdiction reflects systematic variation across jurisdictions in prosecutor and judge behavior, defense attorney quality, and jury preferences. The electorate plays an important role by electing prosecutors and judges, serving as jurors, and by indirectly determining the level of funding for indigent defense. The ϵ_{ict} term reflects idiosyncratic determinants of the charge outcome, including the specifics of the charge, the specific prosecutor, judge, defense attorneys, and if applicable, jury composition.

To identify punishment norms in practice, we first estimate linear models and use the rich observable charge and defendant characteristics included in our data as controls. In section 3.2.3 we explore using case-level characteristics. Specifically, we estimate models of the following form, *separately by state*:

$$Y_{ict} = \tau_{cth(i,t)} + X_i \gamma^X + Z_{it} \gamma^Z + \theta_{j(i,c,t)} + \epsilon_{ict} \tag{2}$$

where i indexes individuals, c indexes initial charge, and t indexes year; $h(i,t)$ is the criminal history for individual i at time t; $j(i,c,t)$ is the court jurisdiction; θ_j are jurisdiction fixed effects ('punishment norms'); $\tau_{cth(i,t)}$ are arrest charge by criminal history by year fixed effects; Y_{ict} is an indicator for any confinement sentence. Lastly, X_i and Z_{it} are vectors of time invariant and time-varying individual characteristics. X_i includes defendant race, ethnicity, and gender; Z_{it} includes age.

To construct criminal history $h(i,t)$ we adopt the federal classification system in all states. This is a point system based on prior offenses. For each prior offense, a defendant receives: 3 points if the sentence was longer than 390 days, 2 points if the sentence was longer than 60 days, and 1 point for a conviction. Defendants are assigned to 1 of 6 categories depending on total prior points.[7] This approach introduces measurement error in that states have their own criminal history classification schemes, but it allows for consistency across states.

Note that we model punishment norms as separable from other charge characteristics. We assess this assumption below.

The coefficient estimates for equation (2) are presented in Table 4. The pattern of coefficients is consistent with past research (for example, Rehavi and Starr, 2014). Conditional on offense charge, criminal history, year, and jurisdiction, black and male defendants are more likely to receive confinement sentences. Where the data are available, Latino defendants are also more likely to receive confinement sentences. The relationship between punishment and defendant age is nonlinear, increasing in age at younger ages and eventually decreasing.

Punishment norm estimates are summarized in Table 5. Notably, controlling for observable offense and defendant characteristics does not substantially mute the cross-jurisdiction variation in confinement rates. The ratio of standard deviations for punishment norms to confinement rates ranges from 0.82 in Virginia to 0.89 in North Carolina. Stated differently, punishment norms explain 67-79% of cross-jurisdiction variation in confinement rates.

Table 5 also includes the (populated-weighted) average adjusted confinement rates for juris-

[7] The six categories consist of defendants with 0-1 points, 2-3 points, 4-6 points, 7-9 points, 10-12 points, and 13 or more points.

Table 4: Coefficient Estimates from Punishment Norm Models

Outcome: Confinement	Alabama	North Carolina	Texas	Virginia
Black	0.043**	0.021**	0.063**	0.022**
	(0.001)	(0.000)	(0.000)	(0.000)
Latino		0.034**	0.055**	
		(0.001)	(0.000)	
Male	0.042**	0.031**	0.085**	0.034**
	(0.001)	(0.003)	(0.001)	(0.000)
Age	0.000**	0.005**	0.009**	
	(0.000)	(0.000)	(0.000)	
$\text{Age}^2 \times 100$	-0.002**	-0.005**	-0.010**	
	(0.000)	(0.000)	(0.000)	
Criminal History × Charge × Year Fixed Effects	✓	✓	✓	✓
Jurisdiction Fixed Effects	✓	✓	✓	✓
N Charges	1,448,900	4,745,191	5,857,566	2,893,958
Adjusted R^2	0.206	0.100	0.206	0.184
Mean Confinement	0.271	0.090	0.402	0.161

Standard errors clustered by defendant in parentheses.
˜ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.

dictions in the 1st and 4th (population-weighted) quartiles of jurisdictions, ranked by punitiveness. We construct adjusted confinement rates using the predicted confinement rates for each jurisdiction based on the overall composition of charges in the state. The differences in confinement rates between quartiles is substantial. Across states, defendants are 2 to 3 times more likely to face a confinement sentence in 4th quartile jurisdictions than 1st quartile jurisdictions.

3.1 Within-Defendant Analysis

In the analysis above we control for rich offense and charge characteristics that should account for a substantial portion of factors other than jurisdiction-specific punishment norms that determine charge outcomes. However, it is possible that we are omitting critical unobservable determinants. For example, we do not have direct measures of defendant socioeconomic status, which may affect outcomes directly or through defense attorney quality. We may also miss unobservable severity of the offense or other characteristics of the defendant (e.g. perceived crime risk) that may have important implications for charge outcomes.

To evaluate sorting on unobservables, we exploit the fact that many defendants are arrested multiple times and in *multiple jurisdictions.* The movement of a defendant from one jurisdiction to another provides a quasi-experiment for validating punishment norm estimates. Within-defendant comparisons net out time invariant defendant characteristics that contribute to charge

Table 5: Summary of Punishment Norms

	Alabama	North Carolina	Texas	Virginia
Confinement Rate (%)	27.2	9.3	40.8	16.5
SD of Punishment Norms	9.9	2.4	13.1	4.0
Number of Jurisdictions	67	100	253	117
Adjusted Q1 Rate	16.2	6.4	24.5	11.2
Adjusted Q4 Rate	42.2	12.4	60.9	22.2

Notes: Statistics weighted by jurisdiction population. Punishment norms estimates are derived by estimating equation 2 separately by state. The outcome is an indicator for any confinement sentence. Further details on the estimation of punishment norms are discussed in section 3.

outcomes, and we can assess the importance of time-varying unobservable factors by exploiting the timing of the defendant's 'move' from one jurisdiction to another.[8] If punishment norm estimates are unbiased estimates of the causal effect of location, then those estimates should provide unbiased predictions for *changes* in confinement rates for a given defendant that moves from one jurisdiction to another. To the best of our knowledge, this is a novel strategy in this literature.

In Appendix Table A1, we compare charge and individual characteristics for 'mover' defendants, those who have been arrested in multiple jurisdictions, versus 'stayer' defendants, those who have only been arrested in one jurisdiction. Among stayer defendants, we also look separately at defendants who have faced multiple cases. Twenty-seven percent to 40% of defendants have multiple cases in our data, accounting for 59% to 74% of charges. Among defendants with multiple cases, 26% to 41% are arrested in multiple jurisdictions, accounting for 19% to 33% of all charges. Movers are more likely to face confinement sentences than all stayers, and more likely to face confinement sentences than stayers with multiple cases in all states but Texas. They are less likely to be black than all stayers and stayers with multiple cases.

For mover defendants and stayer defendants with multiple cases, we also compare pre- and post-move case pairs for movers and sequential pairs of cases for stayers in Table A2, focusing on the main charge. For 37.5% to 48.7% of mover pairs, the main charge is of the same crime type in each case. This range is 40.9% to 67.2% for stayer pairs. For movers, 53.5% to 67.2% of post-move cases are in counties adjacent to the pre-move case.

To implement this validation strategy, we first randomly partition defendants in each state into 10 equal-sized subsamples. For each subsample, we estimate equation (2) using the other 9 subsamples. We do this to estimate punishment norms for each defendant that are not de-

[8] When we refer to defendants 'moving' from one jurisdiction to another, we are referring to changes in the jurisdiction where they are arrested, not necessarily changes in residence.

rived from that defendant's own case outcomes. We use these (subsample-specific) estimates to predict confinement outcomes for mover defendants in the selected subsample. For these mover defendants we compare the actual change in the confinement rate before and after the move to the predicted change, adjusting for offense and criminal history. That is, for a defendant who faces one charge in county A and one charge in county B, we compare the predicted difference in outcomes between the two charges to the actual difference in outcomes between the two charges.

Formally, we take first-differences of equation (2) to model the *change* in charge outcomes for a defendant moving from jurisdiction A to B:

$$\begin{aligned}
\Delta_i Y_{ict} &= \Delta_i \tau_{cth(i,t)} + \Delta_i X_i \gamma^X + \Delta_i Z_{it} \gamma^Z + \Delta_i \theta_{j(i,c,t)} + \Delta_i \epsilon_{ict} \\
&= \Delta_i \tau_{cth(i,t)} + \Delta_i Z_{it} \gamma^Z + \Delta_i \theta_{j(i,c,t)} + \Delta_i \epsilon_{ict} \\
\Delta_i Y_{ict} - \Delta_i \tau_{cth(i,t)} - \Delta_i Z_{it} \gamma^Z &= \Delta \theta_{j(i,c,t)} + \Delta_i \epsilon_{ict}
\end{aligned}$$

For each defendant i we plug in coefficients for $\tau_{cth(i,t)}$ and γ^Z as well as punishment norms θ_j from the model estimated using the 9 subsamples that do not include defendant i and estimate the following model for movers:

$$\underbrace{\Delta Y_{ict} - \Delta \widehat{\tau_{cth(i,t)}} - \Delta Z_{it} \widehat{\gamma^Z}}_{\text{adjusted change in confinement}} = \alpha + \beta \underbrace{\Delta \widehat{\theta_{j(i,t)}}}_{\text{change in punishment norms}} + \xi_{ict} \tag{3}$$

adding a constant term α to allow for systematic prediction error. A β coefficient of 1 indicates that the punishment norm estimates are unbiased.

The identifying assumption that underpins this validation strategy is that 'mover' defendants do not sort across jurisdictions in a manner that relates to: (1) time-varying unobservable defendant-level or jurisdiction-level determinants of charge outcomes or; (2) match effects–interactions between punishment norms and defendant characteristics.

For example, if defendants that move to a particular jurisdiction are also committing increasingly (and unobservably) more severe crimes, then we would mistakenly identify this jurisdiction as punitive. If a jurisdiction is particularly lenient on drug cases but not other cases, and defendants are more likely to commit drug crimes in that jurisdiction, then we would mistakenly identify this jurisdiction as lenient, when in fact it is only lenient for a particular type of case.[9] We test the first assumption below and assess the second assumption in the next section.

In Panel A of Figure 1, we plot these actual changes against predicted changes by state, pooling by origin and destination punishment norm quartile. The data points fall roughly on the 45° line. We estimate a slope of 0.95 and intercept of 0.00. While we formally reject the null hypothesis that the slope is equal to 1 and punishment norms estimates are unbiased, that

[9] There may also be match effects that are specific to movers. For example, some jurisdictions may be more punitive with 'out of town' defendants than long-term residents. However, if punishment norm estimates predict mover confinement rates well, this would suggest this type of match effect is not important empirically.

bias appears to be very limited in magnitude.[10] We also cannot reject *symmetry* for moves to more punitive and less punitive jurisdictions.[11]

This finding has two important implications. First, we can predict within-defendant changes in confinement remarkably well using data on all defendants. This indicates that punishment norms estimated for all defendants are similar to punishment norms estimated for movers. Second, these predictions are accurate for a variety of defendants as defined by their origin and destination jurisdictions.

We test whether defendants sort on time-varying unobservables using a placebo test. In particular, we test for pre-trends in mover defendant confinement rates prior to the defendant's change in jurisdictions. To do this, we focus on confinement rates for defendants that are charged in multiple cases in one jurisdiction, and subsequently in at least one case in a different jurisdiction. As an illustrative example, consider a defendant that faces criminal cases 1 and 2 in county A, and criminal case 3 in county B. If defendants are sorting on time-varying unobservables, we may see pre-trends in punishment *prior* to the defendant's move, conditional on observable case and defendant characteristics. To test for such pre-trends, we can thus check whether the identity of county B predicts the difference in outcomes between cases 1 and 2.[12] If sorting on time-varying unobservables is not a factor, then future moves should not predict changes in outcomes between cases 1 and 2.

In Panel B of Figure 1, we plot within-jurisdiction changes in confinement rates against predicted changes based on future moves. The points roughly fall on the horizontal line at zero, indicating that future moves do not predict earlier changes in confinement rates.

As another validation exercise, we compare punishment norm estimates derived using equation (2) to estimates derived using only within-defendant variation. To estimate punishment norms using only within-defendant variation we estimate models of the form:

$$Y_{ict} = \tau_{cth(i,t)} + \gamma_i + Z_{it}\gamma^Z + \theta_{j(i,c,t)} + \epsilon_{ict} \tag{4}$$

where γ_i are defendant fixed effects.

We compare punishment norm estimates based on equation (2) and equation (4) in Table 6. As reflected in Figure 1 Panel A, the estimates are quite similar. The correlation between estimates within states ranges from 0.81 in Alabama to 0.98 in Texas. The variation is slightly larger for estimates using defendant fixed effects, due at least in part to added measurement error.

---

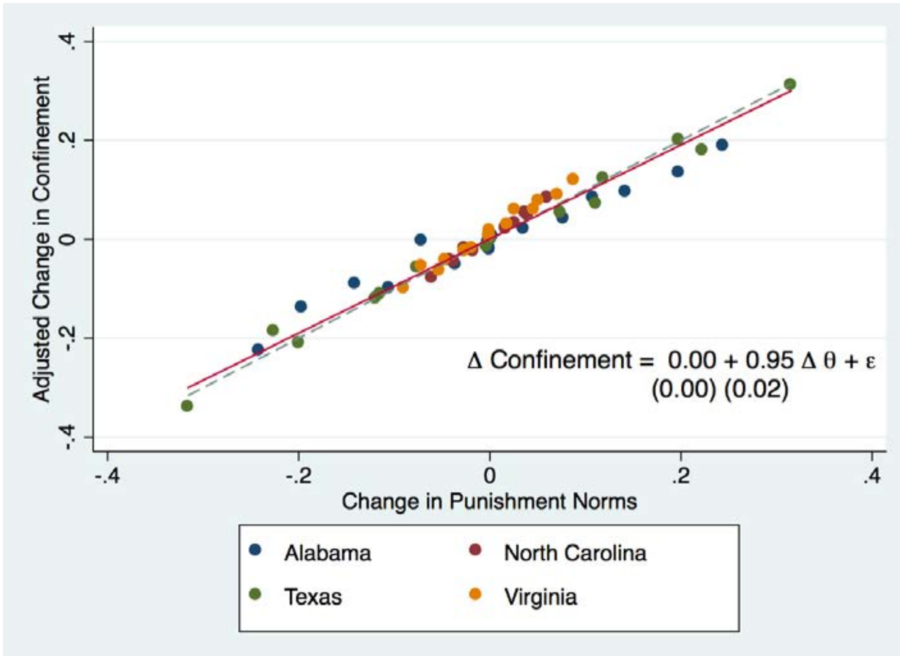
[10] This finding is consistent with past work employing mover-based research designs, i.e. Card et al. (2016).

[11] In particular, if we fit a two-piece linear spline with the knot set at zero, we cannot reject that the two slopes are equal.

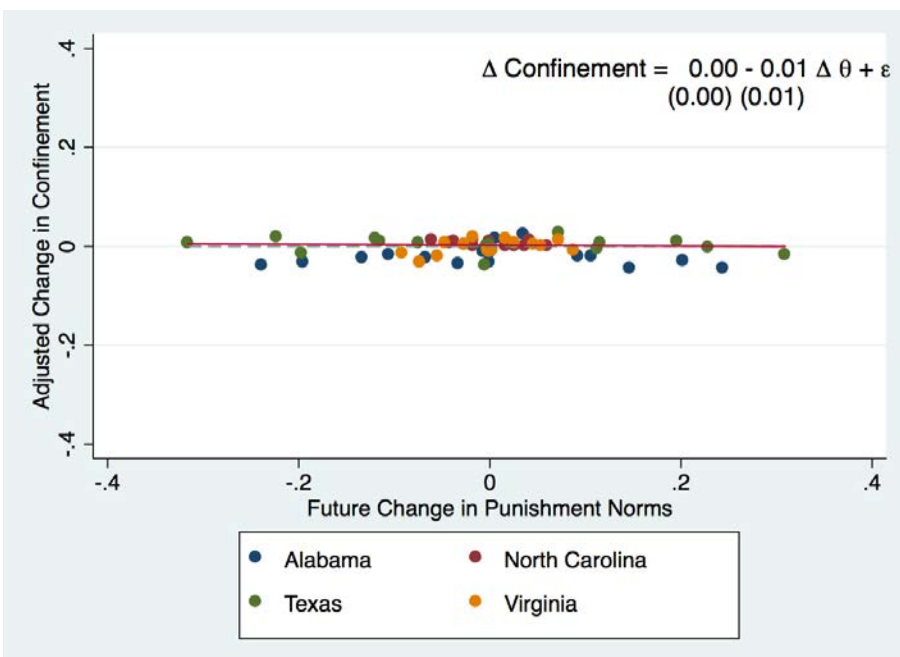
[12] Sorting across jurisdictions based on time-varying unobservables would introduce bias, for example, if defendants that committed increasingly (unobservably) severe crimes were also more likely to relocate to less punitive locations.

Figure 1: Mover Defendant Event Studies

(a) Movers



(b) Placebo



Notes: In Panel A we plot adjusted realized changes in confinement rate before and after the move against predicted changes by state, adjusting for offense and criminal history, and pooling by origin and destination punishment norm quartile. The dashed line is the 45° line, while the solid line is a fitted line through the points, weighted by population. In Panel B we plot within-jurisdiction changes in confinement rates against predicted changes based on future moves. The dashed line is a horizontal line overlapping with the horizontal axis, while the solid line is a fitted line through the points, weighted by population.

Table 6: Summary of Punishment Norms: Overall vs. Within-Defendant

	Alabama	North Carolina	Texas	Virginia
Confinement Rate (%)	27.2	9.3	40.8	16.5
σ (Overall)	9.9	2.4	13.1	4.0
σ (Defendant FE)	10.1	3.0	13.4	5.1
Correlation	0.81	0.92	0.98	0.88
Number of Jurisdictions	67	100	253	117

Notes: Statistics weighted by jurisdiction population. 'Overall' punishment norms estimates are derived by estimating equation 2 separately by state. 'Defendant FE' punishment norms estimates are derived by estimating equation 4 separately by state. The outcome is an indicator for any confinement sentence.

3.2 Robustness Checks

In this section we address three potential confounds for our empirical strategy: (1) match effects; (2) variation in policing; and (3) case-level characteristics.

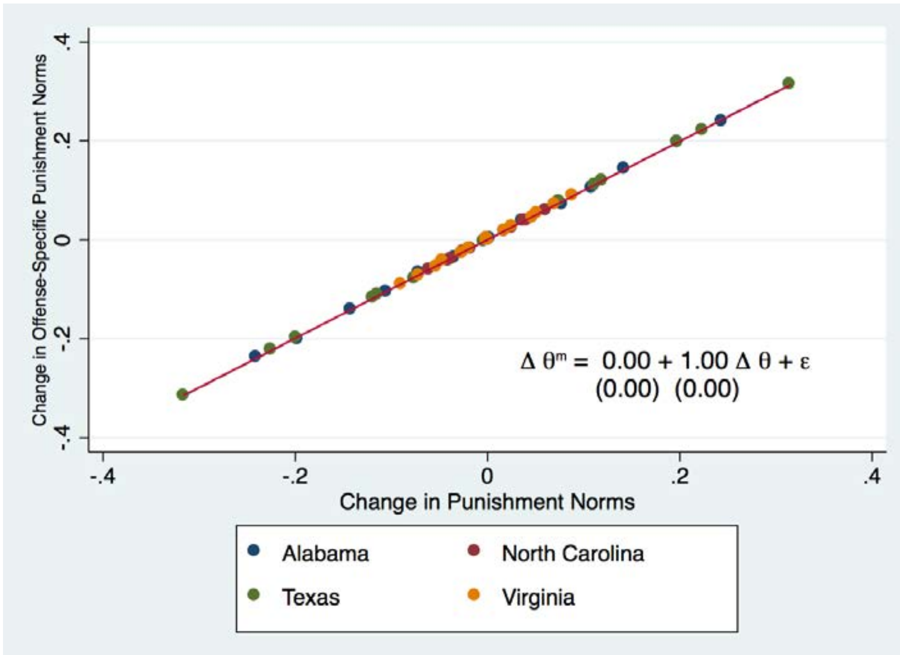
3.2.1 Match Effects

As mentioned above, an identifying assumption for our mover-based validation strategy is that defendant sorting across jurisdictions is unrelated to match effects. To assess this assumption, we first measure the extent to which match effects exist. To do this, we re-estimate punishment norms separately by defendant race (black versus white), by criminal history (first-time versus repeat offenders), and by crime category. We estimate punishment norms separately for property, violent, and drug charges, and for those three 'Core' categories pooled together. We then compare estimates across subsamples.

Population-weighted correlations are presented separately by state in Table 7. The average correlation across race-based severity measures is 0.89, while the corresponding average correlation across criminal history-based measures is 0.92. Punishment norms do not vary significantly by defendant race or criminal history.

The correlation between jurisdiction estimates based on all and core offenses ranges from 0.87 to 0.99. Punishment norms are similar whether or not we restrict to core offenses. The correlations between specific crime categories are generally smaller. The correlations are largest between property and violent crime-based estimates, varying from 0.67 in Alabama to 0.93 in Texas. The correlations between drug crime-based estimates and property and violent crime-based estimates range from 0.50 to 0.96, with most in the 0.5 to 0.7 range. There may be distinct factors that drive punishment norms for drug crimes relative to other crimes. Still, the patterns

Figure 2: Do Defendants Sort on Match Effects?



Notes: We plot changes in jurisdiction by crime type (property, violent, drug, other) match effects against changes in punishment norms before and after the move, pooling by origin and destination punishment norm quartile. The dashed line is the 45° line, while the solid line is a fitted line through the points, weighted by population.

in punishment norms that we describe below are qualitatively similar for charges of each crime type.

Second, we test whether mover defendants appear to sort on jurisdiction by crime type match effects. In particular, we test whether mover defendants that move to jurisdictions with larger estimated punishment norms also commit offenses that are punished particularly harshly or leniently in that jurisdiction. To do this, we first estimate

$$Y_{ict} = \tau_{cth(i,t)} + X_i\gamma^X + Z_{it}\gamma^Z + \theta^M_{j(i,c,t),k(c)} + \epsilon_{ict} \tag{5}$$

using the same procedure described in section 3.1, where $\theta^M_{j(i,c,t),k(c)}$ are crime type match effects, or fixed effects for each jurisdiction by crime type interaction. We then take the sample of mover defendant cases used to construct Figure 1 Panel A, and plot the change in crime type match effects against the change in estimated punishment norms. In the absence of sorting, changes in punishment norms should predict changes in match effects without bias.

The results are depicted in Figure 2. We find no evidence of sorting based on match effects. All the points fall on or very near to the 45° line, and the slope coefficient estimate is 1.

3.2.2 Police Quality

Another remaining measurement concern that could bias cross-jurisdiction comparisons is that the threshold that determines whether an initial charge is filed may vary across jurisdictions. For example, some police departments may be more lenient than others in deciding whether to make an arrest or charge a suspect, or may be worse at collecting evidence that may be required to make an initial charge. In that case, jurisdictions with fewer marginal cases may appear more severe in part because the composition of cases that actually lead to a charge may be (unobservably) more severe.

To assuage this concern, we investigate how a proxy for selection into the court data correlates with estimated punishment norms. Below, we also try to control for this selection when measuring the relationship between punishment norms and jurisdiction characteristics. To proxy for selection, we calculate the ratio of number of charges in the court data for a given county and year to crimes reported in the FBI Uniform Crime Reports (UCR) for the same county and year, and then average that ratio across years by county. We restrict to Part I crimes reported in the UCR data: arson, aggravated assault, burglary, murder, rape, robbery, and theft.

Within states, the correlation between punishment norms and the charge to crime ratio is -0.25 in Alabama, -0.13 in North Carolina, -0.32 in Texas, and -0.22 in Virginia. Jurisdictions that we measure as more punitive also have somewhat fewer recorded charges relative to the number of reported crimes. Reassuringly, when we include the charge to crime ratio as a control variable below, it has little effect on the estimated relationship between punishment norms and jurisdiction characteristics. Moreover, conditional on the characteristics we consider below–population density, in particular–we find no relationship between punishment norms and the charge to crime ratio.

3.2.3 Charges versus Cases

Finally, while we conduct our baseline analysis at the charge-level rather than the case-level for simplicity, this may introduce bias if co-charges contribute to charge outcomes and charge composition within cases varies by jurisdictions.

In case-level specifications, Y_{ict} is an indicator for whether a case results in any confinement sentence. $\tau_{cth(i,t)}$ is defined by the most severe charge faced and the number of additional within-case misdemeanor and felony charges filed against defendant i. We also look at *single charge* cases, where there is no distinction between charge and case.

The coefficient estimates for case-level and single charge versions of equation (2) are presented in Appendix Table A3 and Appendix Table A4. We also correlate our baseline punishment norm estimates with case-level and single charge analogs in Appendix Table A5. Estimates are very similar across approaches, with correlations ranging from 0.89 to 0.99.

Table 7: Summary of Punishment Norms: Subsample Correlations

	Alabama	North Carolina	Texas	Virginia
Correlations:				
Black vs. White	0.96	0.90	0.95	0.90
First vs. Subsequent Offense	0.87	0.90	0.93	0.87
All vs. Core	0.87	0.97	0.99	0.97
Property vs. Violent	0.67	0.78	0.93	0.70
Property vs. Drug	0.69	0.81	0.96	0.63
Violent vs. Drug	0.77	0.61	0.95	0.50
Number of Jurisdictions	67	100	253	117

Notes: Jurisdiction-specific punishment norms are constructed separately within each referenced subsample of defendants/charges. Correlations are then weighted by jurisdiction population.

3.3 Decomposing Punishment Norms: Charging versus Sentencing

Punishment norms may derive from the behavior of several court actors. Prosecutors decide what charges to file; prosecutors and defense attorneys plea bargain; in cases that go to trial, juries decide verdicts; and judges decide sentences. These behaviors are also interdependent. For example, prosecutors may or may not pursue charges depending on the quality of the defense attorney. Prosecutors and defense attorneys plea bargain in the shadow of the judge and potential jury. While we do not have a research design to cleanly separate the roles of each actor type in producing punishment norms, we do have detailed data on the criminal justice process that provide some insight into the relative roles of actor types.

To investigate this question, we divide the criminal justice process into two stages: (1) the decision to prosecute charges and; (2) the final sentence for a charge, conditional on prosecution. The first decision falls primarily under the prosecutor's discretion. The second outcome depends on the plea bargaining between the prosecutor and defense attorney, as well as the judge's discretion.[13].

We re-estimate equation (2) using two alternative charge outcomes: (1) whether the charge is pursued by the prosecution ('charge norms') and; (2) conditional on prosecution, whether the result is a confinement sentence ('sentence norms'). In Table 8, we present the charge rate and confinement sentence rate by state, as well as the standard deviation of charge and sentence norms. We also present correlations between punishment, charge, and sentence norms.

There are three main findings to note. First, variation in charge and sentence norms are

[13] Jury trials for criminal cases are rare. In 2006, 94% of all felony convictions in state courts were the result of a plea agreement (Rosenmerkel et al., 2009).

Table 8: Summary of Punishment Norms: Charging versus Sentencing

	Alabama	North Carolina	Texas	Virginia
Charge Rate (%)	59.6	38.6	77.7	59.1
SD of Charge Norms	9.7	8.3	9.1	6.6
Confinement Sentence Rate (%)	54.6	23.4	51.6	27.3
SD of Sentence Norms	11.7	5.2	14.3	5.5
Correlations:				
Charge vs. Punishment Norms	0.74	0.58	0.53	0.41
Sentence vs. Punishment Norms	0.63	0.59	0.96	0.89
Charge vs. Sentence Norms	0.05	-0.27	0.30	0.01
Number of Jurisdictions	67	100	253	117

Notes: Correlations are weighted by jurisdiction population. 'Charge norms' refers to a re-estimation of equation (2) where the outcome is an indicator for whether the charge is pursued. 'Sentence norms' refers to a re-estimation of equation (2) limited to charges that result in prosecution.

similar in magnitude, and both are similar in magnitude to variation in punishment norms.

Second, both charge and sentence norms are strongly and positively correlated with punishment norms, although the correlations between sentence norms and punishment norms are generally larger. The correlations between charge and punishment norms range from 0.41 in Virginia to 0.74 in Alabama. The correlations between sentence and punishment norms range from 0.59 in North Carolina to 0.96 in Texas.

Third, the correlations between charge and sentence norms vary substantially across states. It is not *a priori* clear what sign to expect. We may expect prosecutors and judges to positively covary across jurisdictions in their punishment preferences, leading to a positive correlation. However, if prosecutors drop the weakest charges, selection may lead to a negative correlation. We find a positive correlation in Texas, a negative correlation in North Carolina, and no correlation in Alabama and Virginia.

We conclude that both charging and sentencing are important contributors to punishment norms. The importance of charge norms suggests that prosecutors play a crucial role in producing local punishment norms.

4 Racial Divisions and Punishment Norms

We have provided evidence in support of a causal interpretation of our punishment norm estimates and established the robustness of our severity measure across alternative outcomes and approaches. We next explore those jurisdiction-level characteristics that predict the magnitude of punishment norms. To guide this analysis, we sketch a simple model of preferences for punishment based on racial group loyalty to derive a predicted relationship between punishment norms and local racial heterogeneity.

4.1 A Simple Model

For the purposes of our model, we assume that local residents have to choose an optimal level of punishment, but are constrained to choose an overall severity level rather than separate severity levels by race.[14] Given this restriction, we model the utility of individual i as follows:

$$u_i(s; p(r_i)) = s \times [\alpha(1 - p(r_i)) + \beta p(r_i)] - c(s) \quad (6)$$

Here, r_i is the racial group of individual i, $p(r_i)$ is the probability that an offender arrested in individual i's home jurisdiction is a member of individual i's racial group, and $c(s)$ is a strictly increasing and convex function (with $c(0) = 0$) characterizing the fiscal and non-pecuniary costs associated with higher severity s.[15] In the expression for individual utility, α and β reflect the relative utility gains associated with punishing outgroup members versus punishing ingroup members (i.e. a negative-valued β implies disutility associated with punishing ingroup members). Based on the existing literature related to racial group loyalty, we make the assumptions that $\alpha > 0$ and $\alpha > \beta$.[16]

To characterize how predicted punishment preferences vary as a function of local racial composition, first consider a jurisdiction in which the share of offenders who are white (p_w) is substantially greater than one-half. In this case, the severity level preferred by white residents, $c'^{-1}(\alpha(1 - p_w) + \beta p_w)$, will be lower than $c'^{-1}(\alpha(p_w) + \beta(1 - p_w))$, the severity level preferred by black residents. Now, suppose that the pivotal (median) voter is the one whose preferences determine the jurisdiction-specific severity level. Since racial population shares are highly correlated with the share of defendants of each race, the likelihood that the pivotal voter is white is increasing in the share of defendants that is white, and so white punishment preferences will determine local severity. Next, note that as the black share of offenders $(1 - p_w)$ increases, the

14 This assumption is justified empirically by the findings that (1) incarceration policy severity in a given jurisdiction is highly correlated across racial groups and (2) there is no clear relationship in our sample between those jurisdiction characteristics that predict overall jurisdiction-level severity and the gap between within-jurisdiction black and white defendant-specific severity parameters. The latter finding is discussed in more depth below.

15 For example, increased punishment s may impose an additional non-pecuniary cost to the extent that an increase in the likelihood of type II errors, whereby innocent individuals are incorrectly punished, decreases utility (due either to fairness concerns or an individual's self-interested concern that he/she may be erroneously convicted of a crime).

16 Luttmer (2001) and Chen and Li (2009) provide observational and experimental support for these assumptions. Anwar et al. (2012) find that all-white jury pools convict black defendants significantly more often than white defendants, and this gap in conviction rates is eliminated when the jury pool includes at least one black member. These findings are consistent with jurors preferring to punish outgroup defendants over ingroup defendants.

severity level preferred by white residents will also increase given that $\alpha > \beta$ and that $c'^{-1}(\cdot)$ is a strictly increasing function by construction. Hence, the severity level chosen by the median voter is increasing in black offender share until the median voter switches from a white to black resident. By the symmetry of the model, the severity level preferred by black residents is falling as the black share of offenders continues to rise. Consequently, the model predicts that local severity levels as a function of the black share of offenders will follow an inverted U-shape.[17]

4.2 Testing the Model

As an initial test of this prediction, Panel A of Figure 3 plots transformed punishment norms for each county as a function of the black share of defendants in that county. We transform punishment norms to make them comparable across states in this and the subsequent analysis. We begin with punishment norm estimates derived from the model (2) using the full data.[18] We then construct the predicted confinement rate for each jurisdiction based on the overall composition of charges in the state using the model estimates. We next divide this predicted confinement rate by the overall state confinement rate and take the log of this ratio. For each jurisdiction, the result is approximately the proportional difference in confinement rates between a jurisdiction and the overall state, holding other charge characteristics fixed. Below we denote this transformed punishment norm $\log \theta'_j$.

The plot reveals that the inverted U-shaped relationship predicted by our model is indeed borne out in the data. For an initial range of values for the black share of defendants, punitiveness is increasing in the black share of defendants. After this range, the sign of the relationship flips.[19]

To clarify this relationship, we pool jurisdictions into bins based on the black share of defendants, where each bin has a range of 5 percentage points (0-5%, 5-10%, 10-15%, and so on). For each bin we then average the adjusted punishment norm described above, weighting by jurisdiction population. The results are presented in Panel B of Figure 3. Note that, for clarity, the span of the vertical access is substantially narrower in this panel. There appears to be an inflection point at a black share of defendants of about 0.4, and the sign of the relationship flips. Note that, if population and defendant shares are equal, voters have uni-dimensional preferences, and all voters are either white or black, the model predicts an inflection point where the black share of offenders/population is equal to one half.

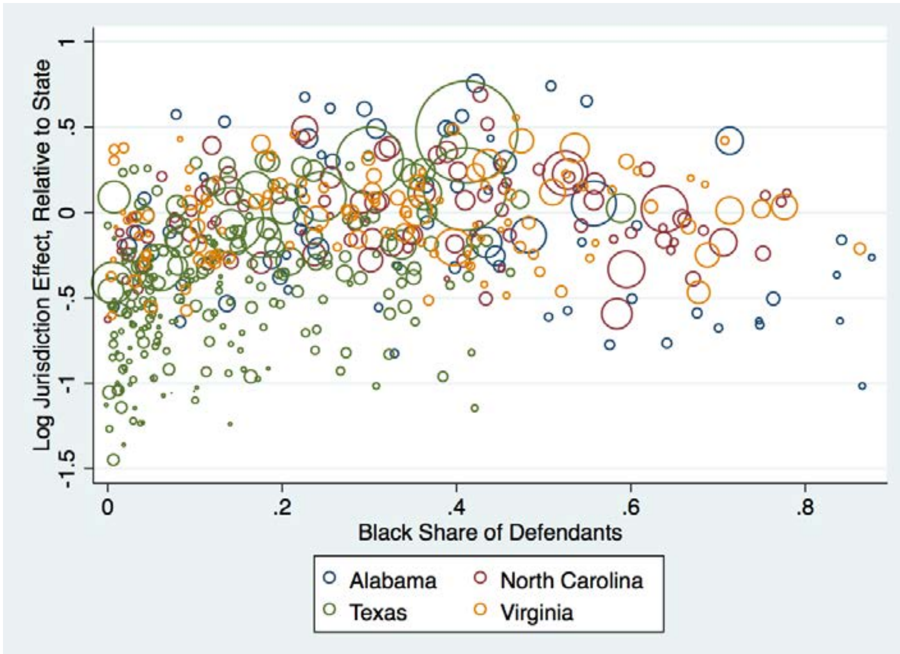
[17] The precise racial population and arrest shares at which we expect to observe this inflection point are uncertain, given variation across jurisdictions in voting rates by race, the share of the population categorized as "Other race", and the multidimensionality of policy preferences.

[18] All results are very similar if we use punishment norms derived from model (4) which includes defendant fixed effects or if we use subgroup based estimates explored in section 3.2.1.

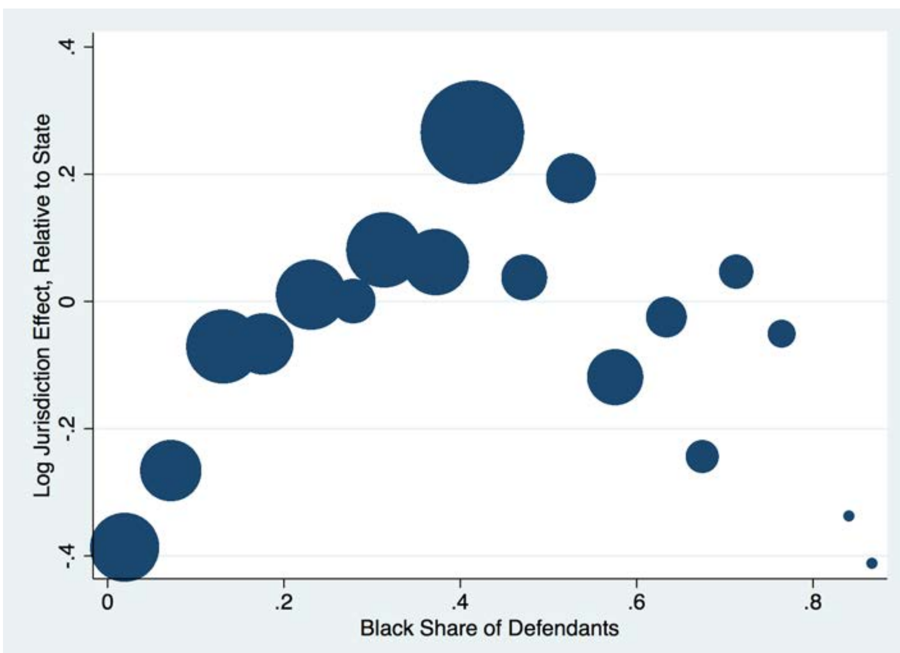
[19] Since punishment norms are estimated with controls for defendant demographics, including race, comparisons across jurisdictions reflect a weighted average of differences in the severity of treatment of black and white offenders (with weights determined by jurisdiction-specific offender shares). This approach eliminates the mechanical relationship between local severity and local black defendant share that would otherwise bias cross-jurisdictional comparisons.

Figure 3: Punishment Norms and Racial Heterogeneity

(a) Raw



(b) Binned



Notes: Marker sizes are proportional to jurisdiction population. Log punishment norms are constructed by dividing the predicted confinement rate for each jurisdiction based on the overall composition of cases within the state by the overall state confinement rate and then taking the log of this value.

We next move to a more formal analysis of the relationship between local severity levels and racial composition. Absent any source of plausibly exogenous cross-sectional variation in racial composition, we introduce a series of additional jurisdiction-level covariates into a regression of log adjusted punishment norms on a quadratic in the black share of defendants to assess the extent to which alternative mechanisms may drive the observed relationship. Specifically, we estimate models of the following form:

$$log\ \theta_j^{'} = X_j\beta + \tau_s + \epsilon_j \tag{7}$$

where $log\ \theta_j^{'}$ are the log adjusted punishment norms described above, X_j is a vector of jurisdiction characteristics, and τ_s is a set of state fixed effects.

Researchers studying US trends in crime and punishment have highlighted the important role that historical violent crime rates played in driving the increased severity of punishment over recent decades and in generating cross-state variation in punishment severity (see, for instance, Western, 2006). To test whether local variation in past crime rates is associated with differences in punishment severity *within* states, Table 9 specifications alternatively include measures of the 1970-1990 violent crime growth rate, 1990 violent crime rates, and 2000 violent crime rates, all measured at the jurisdiction level. Each measure is standardized to have a mean of zero and a standard deviation of one. The crime measures are derived from FBI UCR data. In addition a quadratic in the black share of defendants, we also include log average income, the Gini index of income inequality, Republican vote share in the 2000 Presidential Election, the fraction of prime-aged males in the population, and log population density (measured in 2000). Descriptive statistics for these county characteristics are reported in Appendix Table A9. All specifications weight observations by jurisdiction population, although results are similar if we do not weight by population.

Column (1) presents the regression equivalent of Figure 3. Point estimates are consistent with an inverted U-shaped relationship between local severity and black share of defendants and imply that punishment severity levels are highest in jurisdictions with a black share of defendants equal to 0.43.[20] At this maximum, the predicted value of θ is 54 log points larger than the predicted value where black share is set to zero. This implies that predicted punishment norms are 72% higher in jurisdictions with this level of heterogeneity relative to all-white jurisdictions. This estimated inflection point is quite close to the prediction that punishment severity should be highest in jurisdictions in which one-half of defendants are black. In Appendix Table A10 we use the black share of the *population* in place of the black share of *defendants*. The inverted U-shape

[20] As an alternative approach to testing for an inverted U-shaped relationship between black defendant share and punishment norms, we estimate two piece linear splines and test for a positive initial slope and negative final slope. We estimate a series of splines with knot points ranging from a black defendant share of 0.1 to 0.7. The results are reported in Appendix Table A7, including the slope estimates and adjusted R^2 for each model. Adjusted R^2 is maximized with a knot point of 0.4. In that model the estimated initial and final slope coefficients (standard errors) are 1.31 (0.26) and -1.22 (0.33).

relationship remains, though the implied inflection point moves to 0.30. Across counties, the correlation between the black share of defendants and the black share of the population is 0.92.

Columns (2)-(4) include separately the three measures of historical and contemporaneous crime rates. Results from these specifications lend little support to the hypothesis that within-state variation in present-day severity is explained by lagged local crime rates, crime rate growth, or current crime patterns. Across these three specifications, coefficients on crime are negative and small in magnitude. Column (5) includes the 2000 standardized violent crime rate as well as the additional covariates included in columns (2)-(4), and interacts each of these covariates with state indicator variables. The inverted U-shape relationship between severity levels and black defendant share remains highly significant and the inflection point remains close to one-half in this more flexible specification. Turning to the remaining covariates, population density and Republican vote share both consistently predict higher confinement rates. The positive association between imprisonment levels and Republican political control parallels findings from Western (2006). We also find a positive but imprecise relationship between income inequality and punishment norms (by contrast, Western (2006) estimates a negative but not statistically significant association).

As discussed in section 3.2.2, one concern with interpreting the results in Table 9 is that the type of offenses that lead to charges may vary across counties. For example, jurisdictions with fewer marginal cases may appear more severe in part because the composition of cases that actually lead to a charge may be (unobservably) more severe. To address this concern, we estimate versions of equation (7) that include a jurisdiction's charge to crime ratio as an additional control. The results are presented in Table 10.

We first explore how the charge to crime ratio relates to the covariates we use to explain punishment norms. Columns (1)-(3) report coefficients from linear regression models where the outcome is a jurisdiction's charge to crime ratio. Jurisdictions with higher crime rates and population density have lower charge to crime ratios. Columns (4)-(6) report coefficients for a subset of the models reported in Table 9, adding the charge to crime ratio as a control. Conditional on the jurisdiction covariates we include, the charge to crime ratio is essentially uncorrelated with local severity and the black defendant share.

To assess the validity of the assumption that jurisdiction residents' preferences determine average local severity levels rather than race-specific severity levels, we re-estimate the specifications included in Table 9 in Appendix Table A11 but use the black-white difference in log adjusted local severity as our outcome measure. Across specifications, we see little evidence that race-based gaps follow the same inverted U-shape as the average severity level.

An alternative explanation for the relationship we identify between racial heterogeneity and punishment severity is that (1) a higher share of defendants in racially heterogeneous communities are paired with judges or prosecutors of another race and (2) judges or prosecutors treat outgroup members more severely than ingroup members. Given the paucity of black prosecutors,

Table 9: Punishment Norms and Racial Heterogeneity

Outcome:	Log Jurisdiction Effect, Relative to State				
	(1)	(2)	(3)	(4)	(5)
Black Defendant Share	2.805**	1.878**	1.875**	1.866**	1.628**
	(0.471)	(0.380)	(0.383)	(0.394)	(0.317)
Black Defendant Share,	-3.263**	-2.394**	-2.309**	-2.239**	-1.855**
Squared	(0.547)	(0.512)	(0.508)	(0.550)	(0.480)
Violent Crime Rate Growth,		-0.007			
1970-1990 (Standardized)		(0.017)			
Violent Crime Rate,			-0.033		
1990 (Standardized)			(0.028)		
Violent Crime Rate,				-0.037	x
2000 (Standardized)				(0.025)	
Log Pop. Density		0.089**	0.098**	0.107**	x
		(0.023)	(0.021)	(0.023)	
Gini index		0.321	0.382	0.471~	x
		(0.281)	(0.286)	(0.285)	
Log Avg. HH Income		0.005	-0.041	-0.040	x
		(0.117)	(0.123)	(0.114)	
Fraction Republican		0.451*	0.473*	0.523*	x
		(0.215)	(0.213)	(0.207)	
Fraction Males Aged 15-29		-1.100	-1.348~	-1.131~	x
		(0.674)	(0.743)	(0.679)	
State FEs	✓	✓	✓	✓	✓
Adjusted R^2	0.258	0.385	0.387	0.385	0.449
N	538	538	538	538	538

Notes: Robust standard errors in parentheses. Regressions weighted by jurisdiction population. Fraction Republican reflects the Republican vote share in the 2000 Presidential Election and is constructed using Census data.
'x' denotes inclusion of covariate interacted with state fixed effects.
~ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.

ingroup bias seems unlikely to explain the pattern we observe. In 2014, only 6.6% of chief prosecutors are black in our sample states, and that drops to 2.5% if we exclude Virginia (Reflective Democracy Campaign, 2018). While Shayo and Zussman (2011) document robust evidence of judicial ingroup bias in Israel, findings from the US are mixed and suggest that ingroup bias among judges may be limited. Cohen and Yang (forthcoming) find that among Republican-appointed federal judges, white judges differentially punish black defendants more severely. However, they do not find differential gaps in punishment among Democratic-appointed judges and note that the vast majority of black federal judges are Democratic-appointed. Schanzenbach (2015) finds that federal judges do not exhibit ingroup bias, and Arnold et al. (forthcoming) finds no evidence that racial bias varies with judge race among bail judges in Philadelphia and Miami-Dade counties. While Abrams et al. (2012) finds that black judges impose relatively short sentences on black defendants, they are not less likely to impose confinement sentences on black defendants. Our own finding that the black-white gap in punishment severity does not vary with local racial composition also suggests that judicial ingroup bias is unlikely to explain the relationship between racial heterogeneity and overall punishment severity that we identify. If, for instance, white-majority jurisdictions elected white judges who punished black defendants more severely, we should identify a negative relationship between the black share of the population and the black-white gap in local punishment severity.

5 Simulation

Given evidence that there are significant cross-jurisdictional differences in the severity of criminal punishment and that punishment norms can be interpreted causally, we next simulate the share of charges leading to an incarceration sentence and the race-based gap in this share under a counterfactual in which more punitive jurisdictions adopt the punishment level imposed by communities in their state at the tenth percentile of the predicted confinement rate distribution based on black defendant share. Specifically, within a given state, punishment norms above the predicted level assigned to a jurisdiction at the tenth percentile are adjusted downwards to this level. Table 11 presents a comparison of actual confinement outcomes to the simulated confinement outcomes for whites versus blacks in the four states in our sample. In the simulation, we account for the fact that reduced punishment severity interacts dynamically with our criminal history measure, which is a function of both convictions and time served. In order to do so, we adjust confinement probability to account for the fact that simulated criminal histories will be made shorter than actual criminal histories by the reduction in confinement and conviction rates imposed. Across all four states in the sample, the magnitude of the race-based confinement gap declines when we simulate outcomes. Importantly, this is not a mechanical consequence of the adjusted jurisdiction-specific severity levels. Instead, this finding reflects the fact that black residents of these states disproportionately reside in high-severity jurisdictions. Across states,

Table 10: Punishment Norms and Charges Recorded

Outcome:	Charge to Crime Ratio			Log Jurisdiction Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
Charge to Crime Ratio				-0.010	-0.009	-0.022
(Normalized)				(0.014)	(0.014)	(0.014)
Black Defendant Share	-2.488*	-0.506	-0.087		1.858**	1.620**
	(1.110)	(0.783)	(1.028)		(0.393)	(0.318)
Black Defendant Share,	1.200	1.481	0.550		-2.221**	-1.835**
Squared	(1.288)	(1.004)	(1.437)		(0.549)	(0.479)
Violent Crime Rate,		-0.191**	-0.208**	-0.023	-0.039	x
2000 (Standardized)		(0.057)	(0.069)	(0.025)	(0.026)	
Log Pop. Density		-0.246**	-0.651**	0.119**	0.105**	x
		(0.050)	(0.177)	(0.026)	(0.022)	
Log Avg. HH Income		-0.208	0.414	0.134	-0.041	x
		(0.258)	(0.482)	(0.123)	(0.115)	
Gini index		-0.108	0.687	0.411	0.431	x
		(0.702)	(1.698)	(0.327)	(0.306)	
Fraction Republican		1.433*	-0.681	0.758**	0.538*	x
		(0.571)	(1.129)	(0.223)	(0.210)	
Fraction Males Aged 15-29		-2.462~	-4.341	-0.546	-1.155~	x
		(1.373)	(3.041)	(0.722)	(0.678)	
State FEs	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.080	0.321	0.334	0.311	0.384	0.451
N	536	536	536	536	536	536

Robust standard errors in parentheses. Regressions weighted by jurisdiction population.
'x' denotes inclusion of covariate interacted with state fixed effects.
~ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.

Table 11: Simulation Results

	Alabama	North Carolina	Texas	Virginia
Confinement Sentences per Capita				
White (Actual)	0.0075	0.0048	0.0215	0.0061
Black (Actual)	0.0203	0.0166	0.0293	0.0196
White (Simulation)	0.0056	0.0038	0.0176	0.0049
Black (Simulation)	0.0164	0.0136	0.0240	0.0163
Number of Jurisdictions	65	100	253	117

Notes: Statistics weighted by race-specific jurisdiction population.

the gap in confinement sentences per capita falls by 16-18%, with an average decline of 17%. Correspondingly, the black-specific measure of confinement sentences per capita declines by 17-19%, with an average decline of 18%.

6 Conclusion

We study the role that racial divisions play in explaining the punitiveness of US criminal justice policy by collecting and analyzing administrative criminal justice data from four Southern states. We identify substantial variation across jurisdictions within a given state in the severity of punishment and show that this variation persists even when we include a rich set of charge- and defendant-level covariates. We employ a mover-based identification strategy adopted from the teacher value-added and work-firm wage decomposition literatures and find that unobserved defendant heterogeneity cannot explain the differences in punishment norms that we identify. We find that differences in punishment norms are driven by differences in sentencing conditional on prosecution and by differences in the rate at which charges are pursued by prosecutors. The importance of this latter channel indicates that prosecutorial behavior is an important determinant of local norms. We proceed to write down a simple model of racial group loyalty that predicts an inverse U-shaped relationship between local black share of defendants and punishment severity. This prediction is borne out in the data. Our analysis concludes with a simulation exercise that shows that punishment levels and race-based incarceration rate gaps would decline by 16-19% if more punitive jurisdictions adopted the norms of neighbors that are more racially homogeneous.

While a large literature has documented the connection between racial stratification and support for public goods and redistribution, this research offers novel evidence that racial heterogeneity can be similarly linked to preferences for a 'public bad': more punitive criminal justice policy. Given that blacks are more likely to reside in racially heterogeneous communities in the states in our sample, this finding has important implications for the severity of criminal justice policy faced by the average white versus black resident of these states. Moreover, our findings suggest that large race-based gaps in criminal justice outcomes may persist even in the absence of discriminatory treatment within any given jurisdiction.

References

Abowd, John, Francis Kramarz, and David Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *108.*

Abrams, David, Marianne Bertrand, and Sendhil Mullainathan, "Do Judges Vary in their Treatment of Race?," *The Journal of Legal Studies*, 2012, *41* (2), 347–383.

Alesina, Albert, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 1999, *114* (4).

Alexander, Michelle, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*, New York: The New Press, 2010.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics*, May 2012, *127* (2), 1017–1055.

Arnold, David, Will Dobbie, and Crystal S. Yang, "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, forthcoming.

Berdejó, Carlos and Noam Yuchtman, "Crime, Punishment, and Politics: An Analysis of Political Cycles in Criminal Sentencing," *Review of Economics and Statistics*, July 2013, *95* (3), 741–756.

Card, David, Ana Rute Cardoso, and Patrick Kline, "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women," *Quarterly Journal of Economics*, 2016, *131* (2).

_ , Joerg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 2013, *128.*

Carson, E. Ann, "Prisoners in 2013," September 2014. U.S. Department of Justice, Office of Justice Programs.

Chen, Yan and Sherry Xin Li, "Group Identity and Social Preferences," *American Economic Review*, 2009, *99* (1).

Chetty, Raj, John Friedman, and Jonah Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

Cohen, Alma and Crystal S. Yang, "Judicial Politics and Sentencing Decisions," *American Economic Journal: Economic Policy*, forthcoming.

Collister, Brian, "Texas troopers ticketing Hispanic drivers as white," November 2015.

Dahlberg, Matz, Karin Edmark, and Helene Lundqvist, "Ethnic Diversity and Preferences for Redistribution," *Journal of Political Economy*, 2012, *120* (1).

Dyke, Andrew, "Electoral Cycles in the Administration of Criminal Justice," *Public Choice*, December 2007, *133* (3/4), 417–437.

Enns, Peter K., "The Public's Increasing Punitiveness and Its Influence on Mass Incarceration in the United States," *American Journal of Political Science*, October 2014, *58* (4), 857–872.

Enos, Ryan, "What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior," *American Journal of Political Science*, 2015, *60* (1).

Fischman, Joshua B. and Max M. Schanzenbach, "Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums," *Journal of Empirical Legal Studies*, 2012, *9.*

Glaser, James, "Back to the Black Belt: Racial Environment and White Racial Attitudes in the South," *Journal of Politics*, 1994, *56* (1).

Gottschalk, Marie, *Caught: The Prison State and the Lockdown of American Politics*, Princeton, New Jersey: Princeton University Press, 2015.

Hetey, Rebecca C. and Jennifer L. Eberhardt, "Racial Disparitites in Incarceration Increase Acceptance of Punitive Policies," *Psychological Science*, October 2014, *25* (10), 1949–1954.

Huber, Gregory A. and Sanford C. Gordon, "Accountability and Coercion: Is Justice Blind when It Runs for Office?," *American Journal of Political Science*, April 2004, *48* (2), 247–263.

Keen, Bradley and David Jacobs, "Racial threat, partisan politics, and racial disparities in prison admissions: a panel analysis," *Criminology*, 2009, *47*, 209–238.

Keller, Josh and Adam Pearce, "This Small Indiana County Sends More People to Prison than San Francisco and Durham, N.C., Combined. Why?," https://www.nytimes.com/2016/09/02/upshot/new-geography-of-prisons.html= September 2016. The New York Times.

Key, V.O., *Southern Politics in State and Nation*, New York: Knopf, 1949.

Lim, Claire S. H., “Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges,” *American Economic Review*, June 2013, *103* (4), 1360–1397.

Lim, Claire S.H., James M. Snyder Jr., and David Strömberg, “The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems,” *American Economic Journal: Applied Economics*, October 2015, *7* (4), 103–135.

Luttmer, Erzo F.P., “Group Loyalty and the Taste for Redistribution,” *Journal of Political Economy*, 2001, *109* (3).

Miethe, T.D., “Charging and Plea Bargaining Practices under Determinate Sentencing,” *Journal of Criminal Law and Criminology*, 1987, *78.*

Mueller-Smith, Michael, “The Criminal and Labor Market Impacts of Incarceration,” 2015. Working Paper.

Muhammad, Khalil Gibran, *The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America*, Cambridge, MA: Harvard University Press, 2010.

Nelson, Michael J., “Responsive Justice? Retention Elections, Prosecutors, and Public Opinion,” *Journal of Law and Courts*, Spring 2014, *2* (1), 117–152.

Pfaff, John, “Escaping from the Standard Story: Why the Conventional Wisdom on Prison Growth is Wrong, and Where We Can Go From Here,” *Federal Sentencing Reporter*, April 2014, *26* (4), 265–270.
Reflective Democracy Campaign

Reflective Democracy Campaign, “Justice for All? A Project by the Reflective Democracy Campaign on Who Prosecutes in America,” May 2018.

Rehavi, M. Marit and Sonja B. Starr, “Racial Disparity in Federal Criminal Sentences,” *Journal of Political Economy*, 2014, *122* (6).

Rosenmerkel, Sean, Matthew Durose, and Jr. Donald Farole, “Felony Sentences in State Courts, 2006,” Technical Report, Bureau of Justice Statistics 2009.

Schanzenbach, Max M., “Racial Disparities, Judge Characteristics, and Standards of Review in Sentencing,” *Journal of Institutional and Theoretical Economics*, March 2015, *171* (1), 27–47.

Shayo, Moses and Asaf Zussman, "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics*, August 2011, *126* (3), 1447–1484.

Unnever, James and Francis Cullen, "Reassessing the Racial Divide in Support of Capital Punishment: The Continuing Significance of Race," *Journal of Research in Crime and Delinquency*, 2007, *44* (1).

Unnever, James D. and Francis T. Cullen, "The Social Sources of Americans' Punitiveness: A Test of Competing Models," *Criminology*, February 2010, *48* (1), 99–129.

Walmsley, Roy, "World Prison Population List: Eleventh Edition," 2016. World Prison Brief.

Western, Bruce, *Punishment and Inequality in America*, New York: Russell Sage Foundation, 2006.

A Appendix: Data Description

A.1 Alabama

The data for Alabama are from the Alabama Administrative Office of Courts and shared with us by the Center for Science and Law. The earliest records in the data date back to the early 20th century, though data quality and completeness improves over time. We focus on charges filed between 2000 and 2010. We end in 2010 because in the extract we obtained the share of cases that remain unresolved begins to increase significantly in 2011.

We drop charges with missing data on the defendant, including date of birth, gender, and race. In most of these cases, the defendant listed appears to be an organization (e.g., a bail bond company) rather than a person.

We drop charges with missing dispositions, which appear to generally reflect cases that are on-going. We drop probation violations, appeals, and records that indicate intermediate outcomes, such as the transfer of a case from a lower court to a higher court. We restrict to felony and misdemeanor non-traffic offenses.

To match multiple cases to individuals, we group defendants based on full name and date of birth.

The data include the zip code of the court and a court-specific code, but not the name or the county. We match courts to counties based on the zip code. In ambiguous cases, we manually match charges to counties based on the location of actual courthouses.

In Alabama, criminal cases are handled in Circuit and District Courts. Circuit courts are courts of general jurisdiction, and handle all felony cases. There are 148 Circuit Court judges divided among 41 judicial circuits. District Courts handle misdemeanors. There are 98 judges in 67 District Courts, one court in each county. Each judicial circuit is served by a chief prosecutor ('District Attorney').

Judges for both Circuit and District Courts are elected in partisan elections. The length of term is 6 years. Prosecutors are also elected to 6-year terms in partisan elections. Circuit Court judges and prosecutors are elected at the circuit level. District Court judges are elected at the county level.

A.2 North Carolina

The data for North Carolina are from the North Carolina Administrative Office of Courts. These data contain records for charges initially filed from the 2007 to 2014.

To construct the charge-level data file that is ultimately used in our analysis, we merge case records with offense records (that include disposition and sentence outcomes) based on the unique case identifier provided by the North Carolina Administrative Office of Courts, as well as an identifier for the county in which the charge was adjudicated and the case-specific charge number. We successfully merge 100% of charges to disposition records. While the data do not include unique defendant identifiers, we match multiple cases to the same individual using their full name and address. We use information on the North Carolina structured sentencing offense class associated with a given charge to define a defendant's criminal history at the time that an arrest occurs. We restrict the sample to include only offenses classified as felonies or misdemeanors. Next, we exclude charges for which the same charge is subsequently listed with a final disposition. We also drop charges with intermediate outcomes corresponding to the following recorded dispositions: Superseding Indictment or Superseding Process, Transfer to Superior Court, Probable Cause Found, Change of Venue, and Withdrawn from Superior Court. We drop charges with disposition records that contain missing dispositions, since the structure of the data means that charge dispositions should be available for all included charges. We drop charge-level observations corresponding to probation and parole violations, and we drop observations corresponding to youth aged under 16. Finally, we drop observations that are missing information on defendant age.

To construct our confinement and sentence length outcomes, we convert reported incarceration sentence days, months and years into the number of days sentenced. To do so, we rely on the Minimum Sentence Length values associated with each charge disposition. We categorize a charge as resulting in confinement if (1) a non-zero incarceration sentence is listed and no concurrent probation sentence is listed or (2) the charge results in mandatory confinement based on North Carolina structured sentencing guidelines. To identify charge dispositions for charges with missing sentence records, we rely on the offense file and code a charge as resulting in a conviction if the Convicted Offense Code variable in non-missing (i.e., an offense for which the defendant was convicted is provided). To classify charges as dropped, we construct an indicator variable based on whether the disposition is listed as any of the following: Dismissed by the court, Dismissal by DA, No probable cause, Voluntary dismissal DA, Dismissal with leave by DA, and No true bill returned.

Based on guidance received from the North Carolina Administrative Office of Courts, to identify charges corresponding to a single case, we take the connected set of charges that meet any of the following three criteria: (1) charge records include the same case identifier, (2) one charge has a "consolidated for sentencing" case identifier that matches the case identifier associated with another charge, or (3) charges are filed against the same defendant for the same offense code and on the same offense date.

In North Carolina, criminal cases are handled in Superior and District Courts. Superior Courts handle all felony cases. There are 109 Superior Court judges divided among 50 Superior Court districts. These districts are further grouped into 8 divisions. Every 6 months, elected Superior Court judges rotate from one district to another within their division.[21] District Courts handle misdemeanors. There are 256 judges in 47 judicial districts, one court in each county.

[21] This rotation has occasionally been suspended due to budget constraints.

There are 44 separate prosecutorial districts, each served by one chief prosecutor ('District Attorney').

During the period we study, judges for both Superior and District Courts were elected in non-partisan elections.[22] For Superior Court judges, the length of term is 8 years. District Court judges serve 4-year terms. Prosecutors are also elected to 4-year terms in partisan elections. Judges and prosecutors are elected at the level of their respective districts. Some districts span multiple counties, and some fall *within* a county.

While most Superior Court judges are elected through the process described above, there are also a small number of Special Superior Court judges that are appointed by the governor. As of 2014, there were 12 Special Superior Court judges.

A.3 Texas

The data for Texas are derived from the Texas Computerized Criminal History System (CCH). The CCH is a statewide repository of criminal history data and includes data from various local criminal justice agencies, including arresting agencies, prosecuting agencies, and courts. Agencies are required to report data for all offenses that are Class B misdemeanors or greater. This includes all offenses that would potentially lead to a confinement sentence. The earliest records in the data date back to the early 20th century, though data quality and completeness improves over time. We focus on charges filed between 2000 and 2010.

The structure of the Texas data differs from the data collected from other states in that they are not derived solely from court records. In particular, the data only include court dispositions for offenses that are reported by some arresting agency. In our analysis, we drop offense records with no matched court data. We do this because we cannot code charge disposition in those cases. Of the arrests reported in the data over the years we study, about 85% of arrest records have matched court data. In the extract we obtained, merge rates fall after 2010. A 2011 audit from the Texas State Auditor's Office reports that courts may not submit records because: they encounter an error in the electronic submission process that is not reported back to the court; they lack the state identification numbers of arrest incident numbers required for merging; after an initial submission, they must correct or supply missing information manually via fax, resulting in lower submission rates.

To construct criminal histories for defendants, we use court data dating back to 1996. We stop at 1996 because the rate at which court records are matched to arrest records drops dramatically prior to 1996. Between 1996 and 2000, merge rates with court records are between 65-75%. Results are similar if we instead construct criminal histories using court data beginning in 1985 or 2000.

To measure charge outcomes, we take the original court disposition rather than any subse-

[22] The method of election was changed to partisan elections in 2017.

quent updates (for example, following a probation revocation).

We drop juvenile cases, and all cases for defendants below 16. We also drop cases where defendant demographic information, the offense, or court county are missing.

To match cases across individuals, we use the state identification number provided in the number.

In Texas, criminal cases are handled in District and County Courts. District Courts are courts of general jurisdiction, and handle all felony cases. There are 457 District Courts serving the 254 counties in the state. Each district court corresponds to one judge. Most courts serve a single county. Some courts serve multiple, low-population counties. County Courts handle misdemeanors. There are 508 county courts.

Each county is served by at least one elected chief prosecutor ('County Attorney', 'District Attorney', or 'Criminal District Attorney'). In some counties, felony and misdemeanor cases are led by distinct chief prosecutors. Some prosecutors serve multiple counties.

Judges for both District and County Courts are elected in partisan elections. The length of term is 4 years. Prosecutors are also elected to 4 year terms in partisan elections.

A.4 Virginia

The data for Virginia are derived from administrative records from Virginia's Office of the Executive Secretary.

The data do not include records from Alexandria or Fairfax. This leaves us with data from 118 cities and counties.

We drop charges with missing dispositions, which appear to generally reflect cases that are on-going. We drop probation violations, appeals, and records that indicate intermediate outcomes, such as the transfer of a case from one court to another. We restrict to felony and misdemeanor non-traffic offenses.

To match multiple cases to individuals, we group defendants based on full name and the day and month of birth. The Virginia data exclude year of birth.

In Virginia, criminal cases are handled in Circuit and District Courts. Circuit Courts handle all felony cases. District Courts hear all criminal cases involving misdemeanors. There are Circuit and District Courts in every city. Circuit Courts are divided into 31 circuits. District Courts are divided into 32 districts. Each city and county is also served by one chief prosecutor ('Commonwealth's Attorney).

Circuit Court judges are appointed to 8-year terms by a majority of both houses of the General Assembly. District Court judges are also appointed by the legislature, but to 6-year terms. Prosecutors are elected to 4-year terms via partisan elections.

Table A1: Descriptive Statistics: Stayers versus Movers

State:	Alabama			North Carolina			Texas			Virginia		
	All	Multiple Cases		All	Multiple Cases		All	Multiple Cases		All	Multiple Cases	
		Stayer	Mover		Stayer	Mover		Stayer	Mover		Stayer	Mover
Male	70.2	69.9	74.8	72.6	76.4	77.7	77.2	81.6	81.7	71.8	74.6	77.8
Black	39.7	42.9	26.0	43.8	48.8	36.7	25.1	30.8	22.8	41.3	45.0	39.3
Latino				5.4	3.6	1.8	32.9	35.0	25.4			
Age	33.1	32.4	32.1	32.3	32.1	30.4	31.3	30.4	30.4			
	(11.2)	(10.5)	(9.4)	(12.5)	(12.1)	(10.5)	(11.3)	(10.5)	(9.7)			
Felony				24.8	32.1	32.4	30.9	34.0	32.0	31.9	41.6	46.2
Property	15.9	17.1	20.8	28.0	30.9	39.3	20.9	20.6	24.7	27.0	33.4	38.6
Violent	10.7	8.8	6.6	15.5	15.5	11.8	13.2	13.6	10.0	10.2	9.6	9.3
Drug	17.4	15.7	15.7	16.8	19.2	16.8	21.7	23.8	22.1	13.2	14.9	13.9
Other	56.0	58.4	56.9	39.7	34.5	32.1	44.2	42.0	43.1	49.7	42.1	38.2
Dropped	41.5	37.9	35.9	62.3	63.9	59.4	23.0	20.2	20.7	41.6	41.4	39.0
Convicted	56.3	60.6	62.7	35.1	33.9	38.7	52.8	60.6	60.8	54.1	55.1	57.8
Probation	27.1	31.2	30.8	16.8	15.4	17.1	32.8	25.9	27.2	9.7	10.7	10.4
Confinement	25.6	33.1	33.5	8.1	9.5	11.2	37.8	48.3	45.1	14.0	17.3	21.6
Long Sentence	12.7	16.1	16.4	1.6	2.0	2.0	8.0	10.3	9.9	3.8	5.3	6.3
N Defendants	676,253	146,094	51,166	1,627,960	491,430	211,717	2,161,473	608,108	426,694	1,081,667	193,477	129,322
N Charges	1,504,992	739,123	349,216	3,382,239	2,126,527	1,387,907	3,947,627	2,201,561	1,927,418	2,082,220	888,422	814,835
N Cases	1,010,555	480,396	210,762	2,881,201	1,744,671	1,101,422	3,373,394	1,820,029	1,556,826	1,494,554	598,603	519,442
Charges per Defendant	2.2	5.1	6.8	2.1	4.3	6.6	1.8	3.6	4.5	1.9	4.6	6.3
	(3.6)	(6.0)	(8.5)	(3.2)	(5.0)	(7.2)	(1.8)	(2.6)	(3.3)	(3.3)	(6.5)	(7.1)
Cases per Defendant	1.5	3.3	4.1	1.8	3.6	5.2	1.6	3.0	3.6	1.4	3.1	4.0
	(1.6)	(2.9)	(4.0)	(2.1)	(3.2)	(4.6)	(1.3)	(1.8)	(2.3)	(1.4)	(2.7)	(3.6)

Notes: Standard deviation in parentheses. Missing values reflect outcomes unavailable for particular states. ‘Other’ offenses include crimes against society and offenses we are unable to classify due to miscoding. ‘Long Sentence’ is defined as a confinement sentence of at least 360 days, and is calculated *conditional* on any confinement.

Table A2: Pre- and Post-Move Charges

State:	Alabama		North Carolina		Texas		Virginia	
	Stayers	Movers	Stayers	Movers	Stayers	Movers	Stayers	Movers
Neighboring Counties (%)		67.2		63.8		53.5		64.4
Same Offense Type (%)	67.2	47.8	54.4	38.3	40.9	37.5	60.6	44.4
Pre-Move Charge (%):								
Property	17.1	20.7	33.1	32.6	22.5	26.3	34.4	34.0
Violent	10.3	10.5	15.4	12.1	13.2	9.6	8.6	8.9
Drug	13.8	18.4	18.4	17.9	23.6	22.1	15.3	14.7
Other	58.8	50.4	33.1	37.4	40.8	42.1	41.8	42.4
Post-Move Charge (%):								
Property	16.8	21.4	33.1	32.7	21.7	25.1	32.2	33.1
Violent	9.9	10.9	15.2	12.3	13.8	10.1	8.8	9.6
Drug	13.8	18.6	18.7	18.1	23.0	21.2	15.0	15.0
Other	59.5	49.0	33.1	36.8	41.5	43.6	44.0	42.3
N Case Pairs	285,814	49,870	2,148,180	364,943	1,651,668	676,633	572,927	204,802

Notes: Missing values reflect outcomes unavailable for particular states.

Table A3: Coefficient Estimates from Punishment Norm Models, Case-Level

Outcome: Confinement	Alabama	North Carolina	Texas	Virginia
Black	0.055**	0.025**	0.064**	0.029**
	(0.001)	(0.000)	(0.001)	(0.001)
Latino		0.042**	0.056**	
		(0.001)	(0.001)	
Male	0.051**	0.036**	0.084**	0.037**
	(0.001)	(0.003)	(0.000)	(0.001)
Age	0.000**	0.006**	0.009**	
	(0.000)	(0.000)	(0.000)	
Age^2 $\times$ 100	-0.002**	-0.006**	-0.010**	
	(0.000)	(0.000)	(0.000)	
Criminal History $\times$ Charge $\times$ Year Fixed Effects	✓	✓	✓	✓
Jurisdiction Fixed Effects	✓	✓	✓	✓
N Cases	953,784	3,911,009	4,894,494	1,951,456
Adjusted R^2	0.263	0.147	0.235	0.274
Mean Confinement	0.311	0.111	0.402	0.190

Standard errors clustered by defendant in parentheses.
˜ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.

Table A4: Coefficient Estimates from Punishment Norm Models, Single Charge Cases

Outcome: Confinement	Alabama	North Carolina	Texas	Virginia
Black	0.054**	0.021**	0.063**	0.021**
	(0.001)	(0.000)	(0.001)	(0.001)
Latino		0.039**	0.055**	
		(0.001)	(0.001)	
Male	0.044**	0.031**	0.083**	0.029**
	(0.001)	(0.003)	(0.001)	(0.001)
Age	0.001**	0.005**	0.009**	
	(0.000)	(0.000)	(0.000)	
Age^2 $\times$ 100	-0.002**	-0.005**	-0.010**	
	(0.000)	(0.000)	(0.000)	
Criminal History $\times$ Charge $\times$ Year Fixed Effects	✓	✓	✓	✓
Jurisdiction Fixed Effects	✓	✓	✓	✓
N Cases	671,657	2,979,251	4,236,973	1,501,914
Adjusted R^2	0.277	0.101	0.226	0.238
Mean Confinement	0.277	0.089	0.377	0.150

Standard errors clustered by defendant in parentheses.
~ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.

Table A5: Comparing Punishment Norm Estimates: Charge-Level, Case-Level, and Single Charge Cases

	Alabama	North Carolina	Texas	Virginia
Case-Level:				
Confinement Rate	31.1	11.1	40.2	19.0
SD of Punishment Norms	10.1	3.0	14.0	4.0
Single Charge Cases:				
Confinement Rate	27.7	8.9	37.7	15.0
SD of Punishment Norms	9.9	2.6	13.8	3.6
Correlations:				
Baseline vs. Case-Level	0.916	0.987	0.994	0.944
Baseline vs. Single Charge	0.886	0.957	0.990	0.910
Case-Level vs. Single Charge	0.992	0.975	0.999	0.988
Number of Jurisdictions	75	100	253	118

Notes: Statistics weighted by jurisdiction population. Details on how case-level and single charge case estimates are produced are discussed in section 3.2.3.

Table A7: Testing for a U-Shape with a Linear Spline

Maximum	Adjusted R^2	Initial Slope	Final Slope
0.1	0.181	5.50	0.15
		(1.11)	(0.14)
0.15	0.200	3.77	0.01
		(0.64)	(0.14)
0.2	0.217	2.94	-0.17
		(0.47)	(0.14)
0.25	0.231	2.39	-0.37
		(0.38)	(0.15)
0.3	0.245	1.99	-0.58
		(0.33)	(0.18)
0.35	0.253	1.68	-0.83
		(0.30)	(0.25)
0.4	0.271	1.44	-1.15
		(0.26)	(0.34)
0.45	0.258	1.28	-1.54
		(0.24)	(0.43)
0.5	0.233	1.13	-2.01
		(0.22)	(0.56)
0.55	0.195	0.95	-2.39
		(0.20)	(0.68)
0.6	0.152	0.79	-2.64
		(0.20)	(0.90)
0.65	0.128	0.69	-3.31
		(0.18)	(1.12)
0.7	0.111	0.61	-4.33
		(0.16)	(1.31)

Notes: Each row presents initial and final slope coefficients associated with the corresponding 'Maximum' value, characterizing the black defendant share used to define the row-specific knot point.

Table A8: Arrests and Charges per Reported Crime Across Jurisdictions

	Alabama	North Carolina	Texas	Virginia
Confinement Sentence Per Charge:				
Mean	0.272	0.093	0.408	0.165
SD	(0.115)	(0.027)	(0.147)	(0.049)
Arrests per Crime UCR Part I:				
Mean	0.205	0.276	0.158	0.223
SD	(0.089)	(0.145)	(0.058)	(0.073)
Charges per Crime UCR Part I:				
Mean	0.150	0.432	0.114	0.359
SD	(0.130)	(0.232)	(0.078)	(0.178)
N Jurisdictions	67	100	253	117

Notes: Statistics weighted by county population. 'Confinement Sentence Per Charge' is the rate that *charges* result in a jail or prison sentence. 'Arrests per Crime UCR Part I' is the total number of UCR arrests for UCR Part I offenses (arson, aggravated assault, burglary, murder, rape, robbery, and theft) in a county and year divided by total reported UCR Part I offenses in that county and year, averaged across years. 'Charges per Crime UCR Part I' is the total number of recorded charges for UCR Part I offenses in a county and year divided by total reported UCR Part I offenses in that county and year, averaged across years.

Table A9: Descriptive Statistics for County Characteristics

	Alabama	North Carolina	Texas	Virginia
Black Defendant Share	0.396	0.413	0.250	0.391
	(0.196)	(0.194)	(0.147)	(0.213)
Black Population Share	0.261	0.218	0.118	0.220
	(0.166)	(0.126)	(0.076)	(0.162)
Log Pop. Density	4.969	5.575	6.019	6.077
	(0.989)	(0.985)	(1.613)	(1.622)
Log Average HH Income	9.795	9.899	9.854	9.951
	(0.169)	(0.200)	(0.255)	(0.223)
Gini Index	0.488	0.457	0.510	0.396
	(0.081)	(0.074)	(0.100)	(0.098)
Fraction Republican	0.565	0.560	0.582	0.529
	(0.095)	(0.093)	(0.105)	(0.104)
Fraction Males Aged 15-29	0.104	0.110	0.116	0.106
	(0.015)	(0.025)	(0.020)	(0.030)
Violent Crime Rate,	0.053	1.126	-0.296	-0.545
1990 (Standardized)	(0.674)	(1.239)	(0.595)	(0.805)
Violent Crime Rate,	0.025	0.009	0.153	-0.587
2000 (Standardized)	(1.313)	(1.040)	(0.822)	(1.045)
Violent Crime Rate Growth,	-0.238	-0.371	0.195	-0.002
1970-1990 (Standardized)	(1.018)	(1.165)	(0.853)	(0.690)

Notes: Statistics weighted by county population. Fraction Republican reflects the Republican vote share in the 2000 Presidential Election and is constructed using Census data.

Table A10: Punishment Norms and Racial Heterogeneity in Population

Outcome:	Log Jurisdiction Effect, Relative to State				
	(1)	(2)	(3)	(4)	(5)
Black Population Share	2.718**	1.211**	1.275**	1.353**	1.517**
	(0.670)	(0.407)	(0.435)	(0.424)	(0.382)
Black Population Share,	-4.584**	-1.834**	-1.814*	-1.768*	-2.400**
Squared	(1.075)	(0.687)	(0.704)	(0.720)	(0.660)
Violent Crime Rate Growth,		-0.008			
1970-1990 (Standardized)		(0.019)			
Violent Crime Rate,			-0.037		
1990 (Standardized)			(0.031)		
Violent Crime Rate,				-0.053*	x
2000 (Standardized)				(0.025)	
Log Pop. Density		0.100**	0.107**	0.121**	x
		(0.026)	(0.024)	(0.025)	
Gini index		0.192	0.253	0.399	x
		(0.316)	(0.317)	(0.317)	
Log Avg. HH Income		0.106	0.066	0.063	x
		(0.124)	(0.128)	(0.118)	
Fraction Republican		0.712**	0.718**	0.783**	x
		(0.243)	(0.235)	(0.227)	
Fraction Males Aged 15-29		-0.879	-1.158	-0.930	x
		(0.732)	(0.814)	(0.731)	
State FEs	✓	✓	✓	✓	✓
Adjusted R^2	0.147	0.335	0.340	0.342	0.427
N	538	538	538	538	538

Notes: Robust standard errors in parentheses. Regressions weighted by jurisdiction population. Fraction Republican reflects the Republican vote share in the 2000 Presidential Election and is constructed using Census data.

'x' denotes inclusion of covariate interacted with state fixed effects.

˜ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.

Table A11: Race-Based Confinement Gaps

Outcome:	Black-White Log Jurisdiction Effect		
	(1)	(2)	(3)
Black Defendant Share	-0.417	-0.349	-0.341*
	(0.289)	(0.240)	(0.173)
Black Defendant Share,	0.427	0.375	0.462*
Squared	(0.312)	(0.289)	(0.205)
Log Pop. Density		-0.021	x
		(0.014)	
Log Avg. HH Income		0.042	x
		(0.075)	
Gini index		0.057	x
		(0.151)	
Fraction Republican		-0.032	x
		(0.095)	
Fraction Males Aged 15-29		0.558	x
		(0.346)	
Violent Crime Rate,		-0.001	x
2000 (Standardized)		(0.012)	
State FEs	✓	✓	✓
Adjusted R^2	0.072	0.076	0.100
N	530	530	530

Regressions weighted by jurisdiction population.
'x' denotes inclusion of covariate interacted with state fixed effects.
~ significant at 10 percent level; * significant at 5 percent level; ** significant at 1 percent level.