American Economic Journal: Applied Economics 2015, 7(4): 103–135
http://dx.doi.org/10.1257/app.20140111

The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems[†]

By Claire S. H. Lim, James M. Snyder, Jr., and David Strömberg[*]

We study how media environments interact with political institutions that structure the accountability of public officials. Specifically, we quantify media influence on the behavior of US state court judges. We analyze around 1.5 million criminal sentencing decisions from 1986 to 2006 and new data on the newspaper coverage of 9,828 trial court judges. Since newspaper coverage is endogenous, we use the match between newspaper markets and judicial districts to identify effects. We find that newspaper coverage significantly increases sentence length by nonpartisan elected judges for violent crimes. For partisan elected and appointed judges, there are no significant effects. (*JEL* D72, H76, K41, L82)

A large body of scholarship in political economy asserts that policy outcomes tend to be better when politicians can be held accountable by voters (e.g., see Barro 1973 and Ferejohn 1986). In the absence of such accountability, politicians will use government to advance their own interests rather than those of the electorate. Accountability requires an informed electorate, however. And, in large democracies the media play an essential role in informing voters about the actions of elected officials and therefore in keeping officials accountable.[1]

This paper investigates how the media environment interacts with political institutions that structure selection and incentives of public officials, focusing on judges. Courts play a critical role in securing the stability of an economy by protecting property and enforcing contracts. Thus, understanding the operation of courts is of

* Lim: Department of Economics, Cornell University, 404 Uris Hall, Ithaca, NY 14853 (e-mail: clairelim@cornell.edu); Snyder: Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (e-mail: jsnyder@gov.harvard.edu); Strömberg: Institute for International Economic Studies, Stockholm University, SE-106 91, Stockholm, Sweden (e-mail: david.stromberg@iies.su.se). This paper was previously circulated under the title "Measuring Media Influence on US State Courts." We thank seminar participants at UC-Berkeley, Northwestern-Kellogg School, Stanford GSB, University of Illinois, University of Chicago Harris School, and conference participants at the ASSA annual meeting 2010, CIRPEE Political Economy Workshop, and Erasmus Political Economy Workshop for their suggestions and comments. We also thank Evelina Persson and Christopher Stanton for excellent research assistance. The research conducted for this article was supported by the American Bar Association (ABA) Section of Litigation (Award 05-2010) and the European Research Council (Grant 2106735). The views expressed here are not intended to represent ABA positions or policies.

† Go to http://dx.doi.org/10.1257/app.20140111 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] See, e.g., Thomas Jefferson: "The functionaries of every government have propensities to command at will the liberty and property of their constituents. There is no safe deposit for these but with the people themselves, nor can they be safe with them without information. Where the press is free, and every man able to read, all is safe" (Thomas Jefferson to Charles Yancey, 1816. ME 14:384).

significant importance. Indeed, a large number of studies have been conducted to understand variation in the operation of courts across countries (e.g., see Djankov et al. 2003; Glaeser and Schleifer 2002; and the literature cited therein).

The interaction between the media environment and the selection and incentives of public officials can be studied in detail in the case of US state trial court judges, where a variety of selection systems exist. Some judges are appointed by the governor, some are elected in normal partisan elections after being nominated by political parties, and, most commonly, some are elected in nonpartisan elections where they compete without party identification on the ballot.[2]

State trial court judges exercise enormous power in the US judicial system. State courts handle more than 90 percent of civil and felony cases in the United States. In 2006, they convicted over 1 million felons to a total of over 2 million years in prison (Rosenmerkel, Durose, and Farole 2009). In the state judicial system, while juries have the power to convict, judges have the authority to impose sentences, and only a small fraction of felony cases are reviewed by appellate courts. Consequently, the decisions of trial court judges are of paramount importance, and therefore, so are the selection and incentive structures these judges face.

The media may matter because the citizens who monitor judges—and in most cases also elect them—have little reason to gather information unless they are personally involved with the courts. The vast majority of voters say that they have insufficient information about judicial candidates (Sheldon and Lovrich 1999). The media sometimes provide information about judicial candidates, but often do not. Consequently, monitoring judicial behavior is difficult for voters, and given their weak prior beliefs, even a single news story covering apparent judicial malfeasance can decisively influence elections. As we show empirically, there is tremendous variation in our sample in the amount of newspaper coverage about judges, ranging from no coverage to hundreds of articles per newspaper and year. Consequently, there is similar variation in the ability of voters to effectively monitor the judiciary.

How does this newspaper coverage affect judges' behavior? There is substantial survey evidence that ordinary voters believe that criminal sentences are too lenient.[3] Assuming this to be the case, better monitoring through media coverage is likely to lead to harsher sentencing. Specifically, newspaper coverage may help voters to *select* judges that have harsh penal preferences. It may also *incent* incumbent judges to avoid lenient sentencing. Harsher sentencing could also arise if news stories induce more punitive attitudes, for example, because media are more likely to report underpunishment than overpunishment. Media coverage is likely to have the largest effects where voters have the weakest prior information about judicial candidates (that is, where voters do not know judges' party affiliation) and the most direct

[2] The influence of judicial selection systems on rulings in lawsuits has also been a major public policy concern. For example, a US Supreme Court case, *Caperton v. A.T. Massey Coal Co.*, illustrates how campaigning during elections might create biases in judicial decisions that affect businesses. In the case, a judge presided over a trial in which one of the litigants was a company that provided campaign funds to the judge in his first election. For details, see http://www.supremecourt.gov/opinions/08pdf/08-22.pdf. In response to the public concern, the former US Supreme Court Justice Sandra Day O'Connor has campaigned to remove direct elections of judges. See http://www.nytimes.com/2009/12/24/us/24judges.html.

[3] Section B of the online Appendix presents several examples of lenient sentences that caused controversies and media coverage.

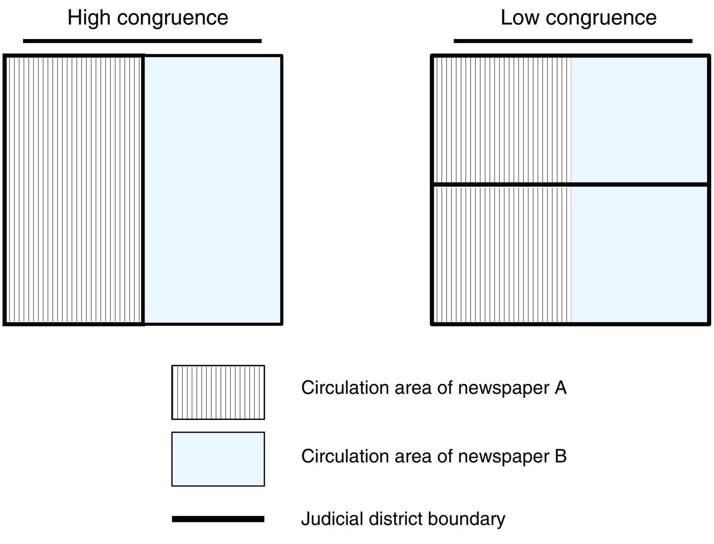

Figure 1. Examples of High and Low Congruence

influence on judge selection (that is, where judges are elected rather than appointed), and for crimes most covered by the media. We assert the following hypotheses: (i) media coverage increases sentence length; (ii) this effect is larger for nonpartisan elected judges than partisan elected or appointed judges; and (iii) the influence of media on sentencing is increasing in the severity of crimes.

We investigate these hypotheses using data on 1.5 million sentences handed down between 1986 and 2006 collected within the National Judicial Reporting Program (NJRP). We combine this with newly collected data on the newspaper coverage of 9,828 trial court judges during 2004 and 2005. We find an average of nine newspaper articles covering each judge each year. We also find that the variation is very large; 1 standard deviation is 21 articles.

Newspaper coverage of judges is most likely endogenous to sentencing harshness. First, severe crimes attract media attention. Second, areas with ample newspaper coverage of courts may have different sentencing outcomes even without newspaper coverage. Consequently, a regression of sentence length on newspaper coverage is unlikely to capture causal effects. To address this concern, we use the match or "congruence" between newspaper markets and judicial districts to identify effects, which we define in a way similar to that in Snyder and Strömberg (2010), a study of newspaper influence on US congressmen. Figure 1 illustrates how the notion of congruence captures variation in newspaper coverage in a manner that reduces concerns about confounding factors.

The left panel of the figure illustrates a case of perfect (high) congruence between newspaper markets and judicial districts. In this case, newspapers cover many stories about their courts (judicial districts) because their courts are relevant to all of their readers. The right panel shows the opposite (low congruence) case, where boundaries of newspaper markets and judicial districts are orthogonal. In this case, whichever court the newspaper chooses to cover in a story, the story is relevant to only half of its readers. Thus, newspapers would cover fewer stories about courts.

We formally define *Congruence* in Section IIIA. We indeed find, in Section IVA, that *Congruence* is a strong predictor for the amount of coverage newspapers have of courts. Further, we find that voters are more likely to regard newspapers as an important source of information in districts with a higher level of *Congruence*.

However, unlike the data by Snyder and Strömberg (2010), which had a source of exogenous cross-time variation by congressional redistricting, our data have a significant variation only in cross section. Thus, *Congruence* may be correlated with the error term. We address this issue by analyzing various channels through which *Congruence* could be correlated with sentencing harshness. Specifically, our preliminary analysis shows that (i) sentencing harshness predicted by observable characteristics of judicial districts is not strongly correlated with *Congruence*, and the estimated relationship is in the wrong direction; (ii) *Congruence* is not correlated with voters' penal preferences; and (iii) *Congruence* is not correlated with distribution of crime types. Our main analysis also shows that regression results of sentencing harshness on *Congruence* is robust to a large set of controls. These results alleviate the concern that unobserved heterogeneity of judicial districts correlated with *Congruence* may cause a spurious relationship between *Congruence* and sentencing harshness.

We find that newspaper coverage significantly increases sentence length. The effects of coverage are sizable. A 1 standard deviation increase in *Congruence*—which translates into 8 more articles per judge per year in the judicial district—is estimated to increase the average sentence length by nonpartisan elected judges for homicides, sexual assaults, and robberies by about 5.7 months (3.4 percent).[4] The estimated effects are significantly lower for appointed and partisan elected judges; in fact, they are not significantly different from zero for either of these two subsamples. We also find that the media effects are monotonically increasing in the ratio of newspaper articles to convictions: highest for the most violent crimes, followed by other violent crimes, property crimes, and drug-related crimes.

Our paper is related to studies on how media affects the elections and policy choices of the most prominent national and state-level politicians (Besley and Burgess 2002; DellaVigna and Kaplan 2007; Dyck, Moss, and Zingales 2008; Enikolopov, Petrova, and Zhuravskaya 2011; Gentzkow 2006; Snyder and Strömberg 2010; Strömberg 2004). While we know something about how the media affect those high-level politicians who are constantly in the media limelight, we know very little about how it affects the myriads of lower-level officials that voters need to monitor.

An important novel feature of our study is that we document media influence on the judiciary, which is a down-ballot office characterized by low levels of voter information and scarce media coverage. A typical newspaper prints around 9 articles per judge per year, compared to about 100 stories per congressman and 1,000 stories about the governor per year. Obviously, media effects are likely to be different in these low information environments. We find that the media effects per newspaper

[4] In our data, the average sentence length in these three categories combined is 166 months. For the average sentence length in each category, see Table A.4 in the online Appendix. To keep the magnitude of the effect (3.4 percent) in perspective, it is useful to compare it to the effect of the criminal history of defendants. In a study of sentencing in Texas by Lim, Silveira, and Snyder (2015), which uses detailed data on criminal history, the authors find that one more violent crime in the criminal history increases sentence lengths by approximately 6 percent.

story are large, but that these effects are easily crowded out by other information cues, such as party labels. In contrast, previous studies on major politicians find strong effects in partisan elections for high-level politicians. Our study has broad implications because many local officials are selected in a manner analogous to the process used for selecting judges. For example, public utility regulators and school board superintendents have similar variation in their selection rules, that is, some states appoint them and others elect them.

Another important contribution of our paper is that we study behavior at the level of the individual politician. As just noted, studies by Besley and Burgess (2002); Eisensee and Strömberg (2007); Snyder and Strömberg (2010); and Strömberg (2004) provide evidence that the media environment matters for public policy outcomes. With one exception, however, these papers only show indirectly that politicians respond to media pressure, because they do not have measures of the decisions actually made by individual politicians. For example, Strömberg (2004) shows that counties in the United States with better access to radio received more federal aid during the New Deal. He hypothesizes that the effects are due to decisions made by state governors. We do not know whether this is the case, however, because many other actors can potentially affect the distribution of funds, including state legislators, local politicians, US representatives and senators, and bureaucrats at all levels. Since the policymaking environment is so complex, it is difficult to pin down the role played by the media.[5] The exception noted above, which directly shows that politicians respond to media pressure, is Snyder and Strömberg (2010). They examine the roll-call voting decisions by members of US Congress, as well as their decisions to appear as witnesses at congressional committee hearings (often a form of "lobbying" for constituents' interests). They find that representatives who are subject to greater attention by local newspapers are less partisan and more likely to appear as witnesses during congressional committee hearings.[6] To better understand the impact the media has on political accountability at the level of the politician, we need more studies of decisions made by individual politicians. One setting where we can observe behavior at the individual level is the judiciary.

Finally, our paper further contributes to the literature on the functioning of judicial selection systems. Hall (2001) and Bonneau and Hall (2009) document statistics of various types of judicial elections, such as the defeat rate of incumbents and the average vote share of winners. Several studies also document the empirical relationship between selection systems and court decisions, e.g., Hanssen (1999, 2000); Huber and Gordon (2004); Gordon and Huber (2007); Lim (2013); and Tabarrok and Helland (1999). Our study enriches the understanding of the selection systems

[5] Similarly, Besley and Burgess (2002) show that areas in India with higher newspaper circulation are more likely to receive aid in response to natural disasters affecting local food production. They note that "elected state governments" are in charge of relief. But who, precisely, in the state government makes the decisions and feels the pressure of the media?

[6] Another paper that employs a relatively direct measure of behavior is Ferraz and Finan (2011), but their findings are even more subtle. They study mayors in Brazil, and show that in areas without access to local media (or local prosecutors), term-limited mayors were significantly more corrupt than nonterm-limited mayors, while in areas with access to local media (or local prosecutors), there was no significant difference in the behavior of term-limited and nonterm-limited mayors. One interpretation of these patterns is that local media prevent the reelection of mayors who are likely to be corrupt.

by uncovering their interaction with the media environment, which (to our knowledge) has not been done.

The remainder of this paper is organized as follows. In the next section, we describe the institutional background of the US state court system and lay out a conceptual framework. In Section II, we describe our data. In Section III, we lay out our empirical strategy and present preliminary analyses that validate it. In Section IV, we present the main results. In Section V, we conclude.

I. Institutional Background and Conceptual Framework

A. *Institutional Background of the US State Court System*

State court systems typically have three layers: state trial court, state appellate court, and state supreme court. State trial courts, which we focus on, are courts of general jurisdiction: they handle general civil and felony crime cases. State trial courts are often called district courts, circuit courts, or superior courts.

Table 1 shows the judicial selection systems used by state trial courts. Currently, there are three major judicial selection mechanisms. The most common is the nonpartisan election system, where multiple candidates compete without party identification on the ballot, and the top two vote-getters compete against each other in general elections (that is, there are runoff elections). In the partisan election system, judicial candidates seek nomination from political parties in primaries, and candidates nominated by parties compete in general elections. Finally, some judges are initially appointed by the governor (or legislature), and when their terms expire they must be reappointed by the governor, or they must run in noncompetitive "retention" elections and be approved by a majority of voters in a yes-or-no vote.

A few states use systems that do not fall into one of the above three categories. For example, in Illinois, New Mexico, and Pennsylvania, judges must run in partisan elections for their initial term, and then run in retention elections for subsequent terms. In three states, New Hampshire, Rhode Island, and Massachusetts, judges are selected by gubernatorial appointment and life-tenured.[7]

This variation in judicial selection systems has emerged over the nation's history. For the first 50 years after US independence, all states appointed their judges; subsequently, partisan elections became increasingly popular, followed by nonpartisan elections. One key driver of judicial reform has been changes in beliefs about the desired degree of judicial independence, and how each system delivers this (Hanssen, 2004a, b). Although many states have changed their selection systems at some point, the time that a state entered the Union is a strong predictor of the type of selection system used today.[8]

[7] We abstract from the difference between appointed judges with life-tenure and those who run for retention elections. Although the two systems may seem quite different, in practice judges rarely fail in retention elections. Hall (2001); Lim (2013); and Lim and Snyder (2015) document that incumbent judges win retention elections more than 99 percent of the time.

[8] Specifically, Hanssen (2004b) argues that whether a state adopts a new system (e.g., changing from the appointment to the partisan, or from the partisan to the nonpartisan) critically depends on how "entrenched" the old system is and that the degree of "entrenchment" can be proxied according to the time the state joined the Union. See figures 4, 5, and 6 in Hanssen (2004b).

Table 1—Selection and Retention Rules for the State Trial Courts

Number of states	Initial selection	Reelection	Set of states
9	Partisan election	Partisan election	AL, IN, KS, LA, MO, NY, TN, TX, WV
22	Nonpartisan election	Nonpartisan election	AR, AZ, CA, FL, GA, ID, IN, KY, MD, MI, MN, MS, MT, NV, NC, ND, OH, OK, OR, SD, WA, WI
3	Partisan election	Retention election	IL, NM, PA
10	Appointment	Retention election	AZ, AK, CO, IA, IN, KS, MO, NE, UT, WY
11	Appointment	Re-appointment or life-tenure	CT, DE, HI, ME, MA, NH, NJ, RI, SC, VA, VT

Note 1: Selection systems can be divided into five groups. Four states (Arizona, Indiana, Kansas, and Missouri) have a within-state variation of two different systems (partisan or nonpartisan election and appointment-retention election) at the district level. These states are included in both categories. For more details, see the website on judicial selection systems by the American Judicature Society (http://www.judicialselection.us/). In New Mexico judges are first appointed by the governor, then they must run in a partisan election, and subsequent elections are retention elections. In Maryland judges are initially appointed by the governor and subsequently run in nonpartisan elections.

Note 2: We classify a state as having nonpartisan elections if party labels do not appear on the general election ballot. In Arizona (in some counties), Maryland, and Ohio, nominations are partisan but the general election ballot is nonpartisan.

Note 3: Arkansas changed its selection system from partisan to nonpartisan, effective in 2002. North Carolina switched from partisan to nonpartisan, effective in 1998. Mississippi switched from partisan to nonpartisan, effective in 1994. These rule changes are reflected in our analysis of sentencing data.

Note 4: Illinois, New Mexico, and Pennsylvania, where partisan-elected judges face retention elections, are classified as the partisan system in our analysis. All the key results in subsequent tables are robust to the exclusion of these three states.

Although judicial elections are down-ballot elections, incumbent reelection rates are not substantially different from those of major elections. Bonneau and Hall (2009) report that reelection rates for state supreme court justices for 1990–2004 are 91.3 percent (94.8 percent for nonpartisan elections), which are comparable to the US House (94.9 percent) and Senate (90.0 percent). Streb, Frederick, and LaFrance (2007) also report that 92.2 percent of incumbents were reelected in intermediate appellate court elections for 2000–2006; 93.4 percent in partisan elections, and 90.6 percent in nonpartisan elections.

B. *Conceptual Framework*

We now discuss how media may influence the courts, and how this depends on the judicial selection system. The media may influence sentencing because it helps voters select candidates whose preferences are aligned well with their own by providing information ("selection effect"). An increase in voter information about candidates may also induce incumbents to avoid decisions that are disliked by voters ("incentive effect"). The media's influence may work through both channels.[9] This is a generic feature in models of media effects on policy makers—see, e.g., section 5 in Prat and Strömberg (2013).[10]

A standard finding in surveys is that ordinary voters want tougher sentencing. For example, the National Annenberg Election Survey interviewed 76,972 US residents living in 2,898 counties in connection with the 2000 Presidential election. The survey asked: "The number of criminals who are not punished enough—is this an extremely serious problem, a serious problem, not too serious, or not a problem at all?" An overwhelming majority (81 percent) responded that this was an extremely serious or serious problem, while only 17 percent answered that this was not too serious or not a problem at all.

Judges who are too lenient in their sentencing often face controversies. For example, in 2013, Judge G. Todd Baugh of Montana sentenced a rapist to only 30 days in jail. An uproar in the community ensued, the case was overturned by the Montana Supreme Court, and the judge was censured. He retired at the end of the term. As another example, in 1991, Judge Joyce Karlin of California gave a nonjail sentence (probation, community service, and a fine) to a defendant who shot and killed a black teenager. The African American community was outraged, and four candidates challenged the judge in the next primary election. For details of these cases and other examples, see Section B of the online Appendix.

Given that voters tend to prefer harsh punishments, more media coverage would help voters identify those judicial candidates that have tougher penal preferences (selection effect) and induce incumbent judges to avoid sentencing too leniently (incentive effect). Media effects of this sort are likely to depend on the judicial selection system.

First, consider the difference between partisan and nonpartisan election systems. In the partisan election system, where the candidate's party affiliation is identified on the ballot, voters may rely heavily on party affiliation, and may not pay the costs to acquire additional information about candidates available from newspapers. Voters might rely on party because it conveys information about a judge's likely

[9] The media may also affect voters' penal preferences, for example, since they carry information regarding how frequently violent crimes are committed or the recidivism rates of convicted criminals. Empirically, we would expect measures of penal preferences to be correlated with media coverage under this channel of influence. We investigate this empirically in Section IIIB.

[10] Several studies on judicial selection systems, without media effects, decompose the selection and incentive effects. Lim (2013) analyzes this question using a structural analysis of sentencing behavior by Kansas trial court judges, and argues that both the selection and reelection incentives affect judges' decisions. Gordon and Huber (2007) also analyze the sentencing behavior of Kansas judges. They argue that partisan-elected judges are harsher than appointed judges. By documenting electoral cycles in the sentencing behavior by partisan-elected judges, they additionally argue that the dominant factor that captures this difference is the reelection incentive. Our data is not well suited to distinguish between these channels since we do not have information regarding which judge handled what case. Consequently, we cannot include judge fixed effects, nor do we know the electoral proximity of the judge sentencing in each case. We explored the effects of electoral proximity using judicial district-level aggregates (e.g., share of judges up for reelection next year) but did not find any significant effects. This may be due to measurement error, resulting in attenuation bias and large standard errors.

From a policy perspective, it is not crucial that we distinguish between selection and incentive effects. Most states that select new judges through nonpartisan (partisan) elections also retain incumbents through nonpartisan (partisan) elections. Likewise, most states that select new judges through appointment also retain incumbents through procedures that differ from those used in nonpartisan or partisan elections. To our knowledge, there has been no serious policy discussion regarding the optimal mixture of different procedures for selection and retention.

criminal sentencing behavior. Or, even if party affiliation is not highly correlated with criminal sentencing behavior, party might convey information about a judge's likely behavior in civil cases (e.g., pro-business, pro-consumer, or pro-labor). Voters might also use party labels because they know that some trial judges become appellate court judges, and that party affiliation is significantly correlated with the behavior of appellate court judges. Alternatively, voters might vote on the basis of their party loyalty simply as a short-cut or tie-breaking rule.[11] In contrast, in the nonpartisan election system, voters do not even know the party affiliation of the judge, have weak priors, and are thus likely to be more strongly affected by any information that they do have available to them. This in turn strengthens both the selection and incentive effects laid out above, in the nonpartisan election system.

A study of voting behavior in judicial elections by Lim and Snyder (2015) finds strong evidence consistent with this view. For example, they document that voting behavior in partisan judicial elections is extremely partisan: the correlation between Democratic vote share in other races and judicial races is over 0.9 in most cases, and in nonpartisan races the correlation is much less than 0.5 in most cases. They also show that more newspaper coverage, captured by *Congruence* as in our study, increases voter turnout in judicial elections only in the nonpartisan system.

Next, consider how media influence may differ between appointed and nonpartisan elected judges. The first possible mechanism of differences is issue-bundling as in Besley and Coate (2003), which works primarily through the selection effect. In the appointment system, voters can influence the (initial) selection of judges only through their selection of the governor who subsequently selects judges. And, judicial appointments or gubernatorial candidates' penal preferences are in most cases relatively minor considerations for voters when electing governors, compared to issues such as taxes, education policy, etc. Thus, even if newspapers provide information about courts, this information will not strongly influence the election of governors or the judges appointed by them.[12]

The second possible mechanism is through the incentive effect. In 11 states with the appointment system (see Table 1), incumbent judges do not have to run for any kind of reelection involving voters. Therefore, media reports about the courts should have little influence on the careers of incumbent judges, reducing the incentive effect. In ten states with the appointment system, incumbent judges must run for "retention" elections at the end of every term, which are noncompetitive. The absence of challengers in retention elections reduces the information available—especially negative information about the incumbent—since challengers have the

[11] On one hand, a number of papers have found significant correlations between party affiliation and decision making for US and state appellate court judges, on a variety of different issues—e.g., Nagel (1961); Goldman (1966, 1975); Tate (1981); and Brace and Hall (1997). See also the large meta-study by Pinnello (1999). On the other hand, Lim and Snyder (2015) and Lim, Silveira, and Snyder (2015) find no significant relationships between party affiliations and sentencing harshness for trial court judges, and Ashenfelter, Eisenberg, and Schwab (1995) find no significant relationships between judges' party affiliations and behavior in civil rights cases. Note also that the informativeness of party affiliation for judges' sentencing behavior may affect the estimated relationship between sentencing harshness and judicial selection systems, but *not* the variation in media influence across systems.

[12] Media coverage of the courts is even unlikely to affect the behavior of voters who care intensely about judicial issues in multi-issue elections (e.g., lawyers, prosecutors, criminals, or victims) because those voters are likely to have sufficient information and strong views independent of media coverage.

strongest incentives to acquire this information. In addition, without a challenger there is no "race" for the media to cover—so, media outlets will tend to ignore retention elections and focus instead on competitive, two-candidate races.[13] To sum up, the appointment system renders little or no incentive effect in the first place. Thus, there is little room for media influence to work through the incentive effect.

Finally, consider the differences between partisan election and the appointment system. This comparison is less clear theoretically than the previous two because both systems have factors that suppress media influence, as discussed above. Thus, this comparison is primarily an empirical question that depends on the quantitative importance of each factor. If party affiliation almost completely suppresses the influence of other factors in the partisan system, and the issue of crime and punishment matters in gubernatorial elections in the appointment system, then the media influence in the partisan system may be smaller than the appointment system. If the media still matter in the partisan system, due, for example, to primary elections in which the party cue is not important, and the issue of crime and punishment is unimportant in gubernatorial elections, then we may observe a larger media influence in the partisan election system.

To sum up, media influence is likely to be greatest in systems with nonpartisan elections for selecting and retaining judges. There is no clear ranking between the appointment and the partisan election systems. Issue-bundling reduces media influence on appointed judges. On the other hand, the availability of party cues in the partisan election system minimizes the influence of other information on judges, thus suppressing media influence.

In Section A of the online Appendix, we present a simple model of media influence on courts that captures some of the key forces discussed above. For simplicity, the model focuses solely on the selection effect. In the model, only a fraction of voters are informed about the incumbent judge's sentencing behavior, and media coverage of judges increases the fraction of informed voters. The model makes several predictions, but the most important for our purposes is that an increase in media coverage has the largest effect on judicial behavior in the nonpartisan election system.[14]

II. Data

A. *Judicial Districts in the State Trial Court System*

In most states, the state trial court is divided into multiple judicial districts. There are approximately 1,700 judicial districts encompassing state trial courts nationwide, with an average population of just under 170,000. Each district typically has multiple judges. On average, there are 6.6 judges per district. Judicial districts

[13] As mentioned in footnote 7, retention elections are essentially rubber stamps. Lim (2013) estimates the reelection probability function of retention elections in Kansas, and shows that it has no relationship to judges' sentencing decisions.

[14] The model we specify in the online Appendix is more general than our conceptual framework above in that the model does not assume that voters have harsher preferences than the judiciary. The model nests such a setting as a special case.

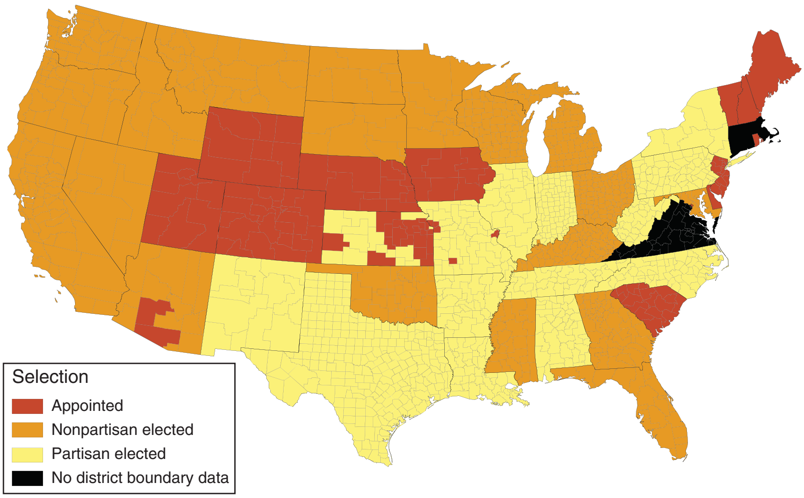

Figure 2. Judicial Districts and Selection Methods

typically consist of a collection of counties.[15] To construct data on the composition of each judicial district over time, we have collected information on the geographic boundaries of these judicial districts for the entire data period, using *The American Bench*.[16] In total, we have data on 1,413 judicial districts.[17] Figure 2 shows a map of judicial districts and selection methods.

B. *Newspaper Coverage*

We collected data on the amount of newspaper coverage regarding state court judges using content analysis. Our sample of judges consists of 9,828 state trial court judges in the United States in 2004 and 2005.[18] We searched 1,186 newspapers for which the articles published in 2004 and 2005 are available through NewsLibrary.com. We analyze 524 newspapers among them, for which we have circulation data from the Audit Bureau of Circulation (ABC). These newspapers comprise 61 percent of the total US newspaper circulation in 2004, as recorded by the ABC. For each judge in our sample and each newspaper with positive sales in the state where the judge presided at the time, we count the number of articles that appeared in

[15] Some small states in New England (e.g., Maine, New Hampshire) have just one judicial district covering the whole state. In the Pacific region (e.g., California) and Mid-Atlantic region (e.g., New Jersey, Pennsylvania), judicial districts tend to cover just one or two counties. In Southern and Midwestern states, judicial districts tend to cover several (three or four) counties.

[16] We first allocated each county to a judicial district using *The American Bench* 2004–2005 edition. To find out if and when each state's judicial district lines were redrawn, we contacted various state officials, typically the director of the administrative office of the judicial branch. We then used the data in the annual series of *The American Bench* to track each such change. We did not collect data on Alaska, Connecticut, Massachusetts, or Virginia, where the county is not the primary geographical unit of the judicial districts.

[17] The number of judicial districts used in our data is smaller than the total number of judicial districts in the nation for the following reasons. First, we exclude Alaska, Virginia, and Massachusetts, in which judicial districts are not completely county-based. Second, for Texas, we use 254 counties rather than 432 judicial districts as the main geographic unit because multiple judicial districts can overlap for the same county.

[18] We obtained the list of judges from *The American Bench*.

Table 2—Summary Statistics

Variable	Observations	Mean	SD	Min.	Max.
Panel A. Newspaper articles and judicial districts					
Total number of articles per judicial district[a]	1,413	122.81	297.43	0	4,566
Number of judges per judicial district[b]	1,413	6.38	17.15	0	389
Number of articles per judge	1,413	16.82	28.40	0	421
Articles per judge per year[c] (market share weighted average)	1,413	9.30	21.35	0	421
Congruence	1,413	0.22	0.30	0	1
Panel B. Sentencing data (homicides, sexual assaults, and robberies)					
Harshness	232,470	0.26	0.34	0.00	1.00
log number of articles	212,837	0.50	0.71	−3.46	2.51
Number of articles	215,750	8.90	16.62	0.00	322.42
Congruence	232,470	0.57	0.21	0.00	0.97
Harsh vote share[d]	166,604	−0.01	0.05	−0.27	0.13
Democratic vote share	230,520	0.57	0.14	0.15	0.90

[a]The unit of observation in panel A is judicial district, not judicial district-year. The total number of articles per judicial district is the sum of all articles covering any judge in a judicial district for the two-year period from 2004 to 2005. The unit of observation in panel B is individual felony case.

[b]Outside of Texas, the minimum number of judges is one. In Texas, we use county as the unit of observation and allocate judges proportionately to populations.

[c]If we divide the total number of articles per judicial district, 122.8, by the average number of judges per judicial district, which is 6.38, we get 19.2 articles per judge for a two-year period or 9.6 articles per judge per year. If we instead use the number of articles per judge per year weighted by *market share* (*circulation*), we get 9.3. It is more correlated with exposure of an individual since people typically only read one newspaper.

[d]*Harsh Vote Share* has a significantly smaller number of observations because ballot propositions are not available in all states.

2004 and 2005 that mention the name of the judge. We use the search string {"judge N1" OR "judge N2"}, where N1 is the judge's full name including middle initial, and N2 is the judge's first and last names only. This yields the frequency of coverage for approximately 1 million judge-newspaper combinations. Since our key variables vary at the judicial district level, we aggregate the frequency of coverage to the judicial district-newspaper level. Summary statistics are shown in panel A of Table 2. On average, a newspaper in our sample publishes nine articles about each judge per year. Coverage varies considerably—the standard deviation of coverage is 21 articles.

A few other comments about coverage are noteworthy. First, to estimate the degree to which coverage of judges focuses on especially violent crime, we ran searches that included the search string {AND (murder* OR rape*)}. In our sample, about 20 percent of the stories contain the added string. Thus, while murder and rape are overrepresented in newspapers relative to the share of criminal acts they represent, they do not dominate the coverage. Second, to estimate the degree to which coverage of judges focuses on their sentencing behavior, we ran searches that included the search string {AND sentenc*}. About 33 percent of the stories contain this added string. Third, inspection of a sample of 200 articles reveals that stories that are not about sentencing cover a wide range of topics, including: election campaigns; candidates' backgrounds, qualifications, and endorsements; election results; judicial procedures and reforms; prison overcrowding and building new prisons or jails; crime rates; laws on the statute of limitations; appellate court rulings; other judicial decisions such as restraining orders; and articles describing ongoing court proceedings in particular high-profile cases. Fourth, we also investigated coverage of courts by local television, using stories in the *Local TV News Media Project*.[19] There appears to be very little coverage of state court judges on local television news. Searching for news stories using the word "judge" yielded just 12 hits, none of which were about sentencing.[20] Searching for the word "sentence," "sentenced," or "sentencing" yielded 35 stories about criminal sentencing decisions or appeals, but none of these mentioned the name of the judge who passed the sentence.

C. *Congruence*

To measure *Congruence*, we use county-level newspaper sales data. Each year, the ABC collects data on each newspaper's circulation in each county for a large fraction of US newspapers. We have this data for 1982 and for the period 1991–2004. For the years 1983–1990 when we do not have circulation data, we interpolate *Congruence*.[21] We complemented this with county-circulation data for non-ABC newspapers for 1991 and 2004, and interpolated values between those years. The non-ABC data is mainly for small newspapers.[22] In our data, the average number of newspaper copies sold in a year is 56 million. The average number of copies sold per household is 0.58, falling from about 0.70 in 1982 to 0.50 in 2004. In Section IIIA, we provide a formal definition of *Congruence* and illustrate its variation with examples from Florida.

D. *Sentencing*

We use felony sentencing data from state courts collected by the NJRP. Felonies are widely defined as crimes having potential punishment of more than one year in prison, and state courts handle the vast majority of felony sentencing (94 percent in 2006).[23] The NJRP collects felony sentencing data from a nationally representative sample of state courts.[24] Data has been collected every two years since 1986 by the

[19] The *Local TV News Media Project*, at the University of Delaware, contains a database with over 10,600 individually digitized stories from over 600 broadcasts from 61 stations in 20 local television markets around the country that aired during the spring of 1998.

[20] One of these stories was about election judges rather than trial or appellate judges, and one was about a judge's funeral, so only ten stories concerned judges' actions or decisions, or judicial elections. Of these, three concerned a judge who was sentenced to jail for fraud, two were about whether a candidate met the residency requirements to run for a judicial office (the candidate was not a sitting judge), one was about a federal judge's decision to strike down Chicago's ban on tobacco and alcohol billboards, one was about a state supreme court's decision that a judge had not violated a state ethics law but had simply exercised free speech, one was about a judge's decision not to quit a trial against tobacco companies, one was about the dismissal of a complaint against a judge for using a racial slur, and one was a retraction by the station of an error in an earlier broadcast.

[21] In Section F of the online Appendix, we check the robustness of our empirical results to dropping the period before 1991.

[22] The non-ABC data was provided by SRDS. On average there are about 10,900 observations each year in the ABC data, and about 500 observations in the non-ABC data. There are about 3,000 counties in the United States, so the average number of observations per county in each year is slightly less than four.

[23] See Rosenmerkel, Durose, and Farole (2009).

[24] The data has been collected through two-stage stratified clustered sampling. In the first stage, counties are divided into multiple strata by their size and the number of felony convictions. Since large counties constitute a large amount of serious crimes, they are given a higher probability of being selected than small counties. At the second stage, a systematic sample of each offense category was selected from each county's records. Rates at which cases were sampled vary by stratum and offense category. In large counties, all murder and rape cases were typically

Census Bureau. For the 1986 NJRP survey, a sample of 100 counties was drawn. New samples of around 300 counties were drawn in 1988, 1996, and 2002.

Our main analysis focuses on 232,470 sentences for homicide, sexual assault, and robbery (around 10,000 sentences every 2 years from 1986 to 1994 and around 30,000 sentences every 2 years from 1996 to 2006). Our unit of observation is individual felony case. The data includes jail time sentenced, as well as offense category and penal codes applied, and demographic characteristics of offenders such as age, race, and gender. The number of judicial districts per year included in this analysis is: 80, 195, 215, 226, 234, 254, 264, 259, 247, 246, and 250 for years 1988–2006, respectively. In Section C of the online Appendix, we show the number of the NJRP sentences by judicial district in Figure A.1, the distribution of judicial districts across states in Table A.3, and the number of sentences and the mean sentence length by offense category in Table A.4.

We use a normalized measure of sentencing harshness relative to other sentences in the same state and year and with the same penal code citation.[25] Once a felon is convicted under a certain penal code citation, it is typically the judge who determines the sentence. Our measure is supposed to capture the discretionary part of sentencing by judges. To construct this measure, we first generate a variable, penal code, which takes the same value for all crimes in each state each year that have the same penal code citation for the first, second, and third most serious offenses. We then identify the minimum and maximum sentence given for that penal code. The variable *Harshness* is defined as

$$Harshness = \frac{sentence - minimum}{maximum - minimum}.$$

This variable is bounded between zero and one, where one means that the judge imposed the highest sentence for this penal code citation in this state and year, and zero means that the lowest sentence was imposed. We code life and death sentences as being of maximum harshness, $Harshness = 1$.[26] Panel B of Table 2 shows the summary statistics of the sentencing data. *Harshness* has a mean of 0.26 and a standard deviation of 0.34. Thus, there is a substantial variation in *Harshness*, but more sentences are closer to the minimum than the maximum harshness.

included, but other offenses were sampled. In small counties, all felony cases were taken. For details, see https://www.ncjrs.gov/pdffiles1/bjs/145323.pdf. Since many counties are repeatedly included in the sample, the combined dataset has an unbalanced panel structure.

[25] Penal code citation conveys much more detailed information than conventional crime (offense) categories, such as NJRP offense categories in Table A.4 in the online Appendix. For example, within the offense category "homicide" there exist separate penal codes for murder in the first degree, in the second degree, manslaughter in the first degree, in the second degree, aggravated murder, etc. Therefore, measuring harshness within cases that have identical penal code citations minimizes the influence of unobserved heterogeneity of felony cases in measurement of sentencing behavior.

[26] We set the maximum sentence to 1,200 months when computing harshness for other sentences in the same state and year and with the same penal code citation as a death or life sentence. This sets harshness below 1 for these sentences, since the maximum sentence is 1,199 months.

E. *Local Penal Attitudes and Controls*

We use two measures of voters' penal preferences. One is the share of voters who vote for the Democratic Party in a presidential election.[27] This measure reflects the general liberalness of voters and is negatively related to the harshness of penal preferences. The other is the share of voters who vote for harsher crime punishment on various ballot propositions. Specifically, we use all available statewide ballot propositions that deal mainly with the punishment of criminals, the rights of the accused, and victim's rights.[28] In virtually all cases, a large majority of voters voted for an increase in harshness toward criminals or the accused, or in favor of victim's rights. On average, more than 65 percent of voters took the harsher position. This is consistent with the widespread view that most Americans believe the criminal justice system is too lenient.

We collected county-level voting data on these ballot propositions from states' election websites and/or election officials. We code all propositions so that higher vote shares represent greater support for increased harshness toward criminals or the accused. For states with more than one proposition, we average the vote shares across the available propositions. We then de-mean the vote shares so that in each state the mean score is zero. We call the resulting variable *Harsh Vote Share*.

There are no surveys that ask respondents about their views on sentencing and are large enough to obtain accurate measures of the average support for harsher sentencing at the judicial district level. The 2000 National Annenberg Election Survey (NAES) probably comes closest to the ideal. As noted above, the NAES contains nearly 77,000 respondents, and contains the following item: "The number of criminals who are not punished enough—is this an extremely serious problem, a serious problem, not too serious, or not a problem at all?" We coded the responses as follows: "not a problem" = 1, "not too serious" = 2, "serious" = 3, and "extremely serious" = 4. We then averaged the scores across districts—call this average *NAES Harshness*. Focusing attention on districts with at least 75 respondents (there are 196 such districts), we find two encouraging correlations. First, the correlation between *NAES Harshness* and *Congruence*, defined in Section IIIA, is quite low, just 0.12. Second, the correlation between *NAES Harshness* and *Harsh Vote Share* is relatively high, 0.61. This gives us added confidence that *Harsh Vote Share* is a reasonable proxy for voters' penal preferences.[29]

We also use data on a number of demographic characteristics at the judicial district level. These have been aggregated from the county level, using data from the US Census Bureau.

[27] For the years without a presidential election, we use linear interpolation.

[28] These propositions are listed in Tables A.5–A.6 in the online Appendix.

[29] One remaining concern in using *Harsh Vote Share* is that only 29 states in our data have ballot propositions as described above. This could cause an endogeneity issue in that states with ballot propositions may have more heterogeneous voters (thus stronger disagreements among voters on the issues of crime and punishment). However, we find no evidence to support this in the data. For example, variation in Democratic vote share in the presidential election, which we can regard as a measure of political heterogeneity across judicial districts, is not larger in states with ballot propositions. Its standard devision is 0.11 in states with ballot propositions, while it is 0.13 in states without them.

III. Empirical Strategy and Preliminary Analysis

A. *Empirical Strategy*

A key concern in identifying the causal effect of newspaper coverage on sentencing is that both may be driven by the seriousness of the crime or characteristics of judicial districts. We use a measure based on newspaper sales called *Congruence* to capture the intensity of newspaper coverage of the courts. We will argue that this measure is exogenous to sentencing.

Congruence measures the match between the newspaper market and the judicial district. When this match is better, in the sense that each newspaper has most of its sales in one district, then newspapers cover the judicial district more. Formally,

$$Congruence_d = \sum_{m=1}^{M} MarketShare_{md} ReaderShare_{md}, \tag{1}$$

where the $ReaderShare_{md}$ is the share of newspaper m's sales that are in district d, and $MarketShare_{md}$ is newspaper m's share of total newspaper sales in district d.[30] The logic behind this measure is that the larger the share of a newspaper's readers that live in a judicial district, the more likely is the newspaper to cover sentencing in that district. The influence of different newspapers is proportional to their market share in the district.

We use variation in $Congruence_d$ to identify effects of newspaper coverage of judges. Note that since *Congruence* is defined using market shares, it is not dependent on the total newspaper penetration in the judicial district. This is important since total newspaper readership in an area tends to be correlated with characteristics such as education, income levels, and interest in politics.

Figure 3 shows *Congruence*, with a special focus on Florida. In some judicial districts, *Congruence* is essentially determined by the readership of a main paper. For example, the judicial district around Jacksonville is highly congruent because the *Florida Times-Union* with a 91 percent market share has 76 percent of its readers within the district. In contrast, the judicial district covering Orlando is less congruent. The *Orlando Sentinel*, its main paper, has a 74 percent market share, but more than half of its readers reside outside of the district.

In other places, a mix of papers is an important determinant of *Congruence*. Consider the three judicial districts at the southeastern corner of Florida, covering Miami, Fort Lauderdale, and West Palm Beach. We call these the South, Central, and North judicial districts.[31] This area has three dominant newspapers, *The Miami Herald* (with 68 percent of its readers in the South and 28 percent of its readers in the Central district), *The South Florida Sentinel-Sun* (with 68 percent of its readers

[30] That is, when we compute *ReaderShare*, we fix the newspaper and compute the weight of each district. And, when we compute *MarketShare*, we fix the district and compute the weight of each newspaper. The notion of *Congruence* is primarily based on *ReaderShare*. Since there are multiple newspapers sold in a given district, we aggregate *ReaderShare* across newspapers using *MarketShare* as the weight of each newspaper.

[31] To be concrete, we discuss the newspaper market in 2006, although the graph covers the average across sample years.

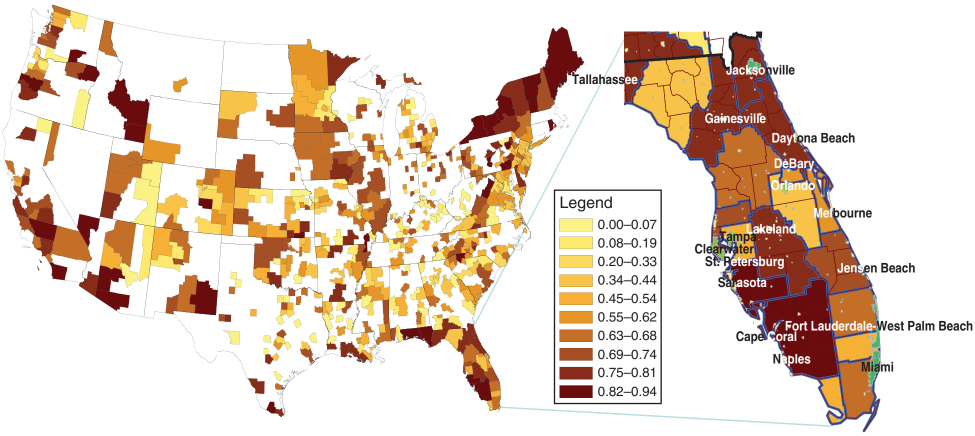

FIGURE 3. *CONGRUENCE*

in the Central and 30 percent of its readers in the North district), and the *Palm Beach Post* (with 86 percent of its readers in the North district). There are also two smaller papers, *The Nuevo Herald* with 88 percent of its readers in the South district and *Boca Raton News* with all of its readers in the North district.

The South district is more congruent than the Central because the second largest paper there, *The Nuevo Herald*, with 27 percent of the market, sells almost exclusively in that district, whereas the second largest paper in the Central district, *The Miami Herald*, with 26 percent of the market, has most of its sales in the South district. The North judicial district is more congruent than the Central because readers of the main paper there (the *Palm Beach Post*) are more likely to reside within the district (and the *Boca Raton News*, with 8 percent of the market, has all of its readers there). *Congruence* for 2006 was 69 percent for the South district, 65 percent for the North district, and 51 percent for the Central district.

In our main analysis, we regress *Harshness* on *Congruence* using a specification of the form

$$(2) \qquad Harshness_{it} = \alpha Congruence_{ct} + \beta_1' \mathbf{d}_{st} + \beta_2' \mathbf{x}_{it} + \varepsilon_{it},$$

where $Harshness_{it}$ is our sentencing harshness measure for sentence i at time t, $Congruence_{ct}$ is *Congruence* in judicial district c that delivered the sentence at time t, $\mathbf{d}_{st}$ is a vector of state-by-year dummy variables, and $\mathbf{x}_{it}$ contains our control variables. We also estimate this equation by judicial selection system. We use *Congruence* as an independent variable rather than an instrumental variable for the number of articles, because we have *Congruence* measure for most of the years in the data period, while we have the number of articles only for 2004 and 2005.[32]

[32] In Section E of the online Appendix, we present an instrumental variable regression version of this specification, by either filling up other years with the data on coverage from 2004 and 2005 or using only the data from 2004 and 2005.

A remaining concern is that *Congruence* might be related to sentencing harshness for reasons other than newspaper coverage. For example, *Congruence* tends to be higher in densely populated areas. Since densely populated areas tend to have a high crime rate, which is also correlated with sentencing harshness, we control for population and area size (in logs). We sometimes use a trimmed sample in which the probability that *Congruence* is above the median is larger than 10 percent and smaller than 90 percent based on population and area size. We also control for other characteristics of judicial districts, but omitted variable bias remains a concern.

One way to mitigate this problem is to use temporal variation as in Snyder and Strömberg (2010). We cannot follow this strategy, however, because judicial district boundaries are rarely redrawn, and newspaper entries or exits are too few to generate sufficient variation to precisely identify *Congruence* effects. Thus, we are forced to rely primarily on variation across judicial districts, and we must worry more about the correlation between *Congruence* and other variables.

To address this issue, we present three preliminary analyses in Section IIIB. First, we investigate the relationship between *Congruence* and variation of *Harshness* predicted by other observables. Second, we investigate whether *Congruence* is correlated with voters' penal preferences. Third, we investigate whether *Congruence* is correlated with distribution of crime types. The relationships between *Congruence* and key demographic characteristics of judicial districts as well as their summary statistics by judicial selection system are documented in Section E of the online Appendix.

B. *Preliminary Analysis*

We first analyze whether the variation of *Harshness* predicted by observable characteristics of judicial districts is correlated with *Congruence*, and whether that correlation depends on the judicial selection system. We conduct this analysis to infer the degree of selection on *unobservable* characteristics of judicial districts using the estimated degree of selection on *observable* characteristics. That is, our strategy is to infer that the correlation between *Congruence* and *unobserved* heterogeneity that affects *Harshness* is likely to be small if our estimated correlation between *Congruence* and *observed* heterogeneity that affects *Harshness* is small.[33] Likewise, if the correlation between *Congruence* and observed heterogeneity does not differ significantly across selection systems, then we can infer that the correlation between *Congruence* and unobserved heterogeneity is likely to be independent of selection systems. This would in turn make our inference on the *difference* in media influence between systems more convincing.

To this end, we first obtain predicted value of *Harshness*, $\widehat{Harshness}$, using the following large set of demographic and political characteristics of judicial districts and case characteristics, identical to the set of variables we use in the main analysis in Section IV: (i) fixed effects for type of crime; (ii) defendant characteristics—

[33] Inferring the degree of selection on unobservables using selection on observables is a conventional strategy used in many empirical analyses in the absence of perfect instrumental variables. For a formal discussion of this strategy, see Altonji, Elder, and Taber (2005).

gender, race, age, and age squared; (iii) crime-related district characteristics—the log total number of convictions in the district, the share of convictions that involved violent crimes, the share that involved drug related crimes, the log total number of crimes in the district reported to the police, and the share of those crimes that were violent; (iv) other district characteristics—log population, log area size, log per capita income, log employment, the share of people in the district who are religious adherents, female, younger than 20, older than 65, black, white, Hispanic, urban, the share with high school education, the share with more than high school education, turnout in the most recent presidential election, and total newspaper penetration; and (v) the interaction between all of these variables and judicial selection systems. We exclude the media variables and the fixed effects for state-and-year. $\widehat{Harshness}_{it}$ we obtain is identical to the predicted value of $\beta_2' \mathbf{x}_{it}$ in equation (2). Then, we estimate

$$\text{(3)} \qquad \widehat{Harshness}_{it} = \delta_1 Congruence_{ct} + \delta_2 d_c^A Congruence_{ct} + \delta_3 d_c^P Congruence_{ct} + \delta_4' \mathbf{d}_{st} + u_{it},$$

where d_c^A is the dummy variable for the appointment system, d_c^P is the dummy variable for the partisan election system, and $\mathbf{d}_{st}$ is a vector of state-by-year dummy variables.

Table 3 shows the results. To understand the extent to which the usage of *Congruence* in place of the amount of coverage mitigates the endogeneity issue, we use the log *Number of Articles* about the judges in a district, instead of *Congruence*, in columns 1 and 2, and *Congruence* in columns 3 and 4. Columns 1 and 2 show that the log *Number of Articles* is significantly correlated with predicted *Harshness* in the nonpartisan system, which suggests a significant potential influence of confounding factors in estimating the influence of newspaper coverage. The correlation also differs significantly across selection systems, which causes a serious concern in analyzing the interaction between media influence and judicial selections systems. In contrast, the correlation between *Congruence* and predicted *Harshness* in the nonpartisan system is statistically insignificant at the 5 percent level, and it is negative. Thus, the usage of *Congruence* significantly alleviates the concern for the influence of confounding factors. Further, the confounding factors are likely to cause a downward rather than an upward bias, if any. The difference across selection systems in the correlation is also statistically insignificant.

Now, we investigate whether voters in different media environments tend to have different penal preferences. The media's focus on violent crimes, for example, may induce a belief that these crimes are more prevalent than they actually are, and that harsher sentences are appropriate. We regressed the share voting for harsher sentencing measures in ballot propositions—*Harsh Vote Share*—on *Congruence*. Table 4 shows the results for all samples and by judicial selection system. We find no significant relationship between *Harsh Vote Share* and *Congruence*.[34]

[34] *Harsh Vote Share* is strongly related to number of crimes known to police, log population, share urban, log employment, and Democratic vote share in the presidential election. We also ran the same regressions with log *Number of Articles* in place of *Congruence* and found no significant relationship.

TABLE 3—REGRESSION OF PREDICTED *Harshness* ON MEDIA VARIABLES, SELECTION SYSTEMS, AND THEIR INTERACTIONS

	Media variable used			
	log *Number of Articles*		*Congruence*	
	(1)	(2)	(3)	(4)
Media variable	0.014***	0.019***	−0.055*	0.005
	(0.003)	(0.003)	(0.032)	(0.027)
Media variable × Appointed	−0.022***	−0.029***	0.050	−0.045
	(0.007)	(0.008)	(0.040)	(0.041)
Media variable × Partisan elected	−0.010**	−0.011**	0.042	−0.052
	(0.004)	(0.005)	(0.035)	(0.033)
Observations	147,497	142,363	163,551	100,983
R^2	0.988	0.997	0.992	0.994
Trimmed sample	No	Yes	No	Yes

Notes: This table shows OLS regression results. The unit of observation is an individual felony case. Standard errors, clustered by judicial district, are in parentheses. All specifications include state-by-year fixed effects. The following set of variables was used to predict *Harshness*: (i) fixed effects for type of crime; (ii) defendant characteristics—gender, race, age, and age squared; (iii) crime-related district characteristics—log total number of convictions in the district, share of convictions that involved violent crimes, share that involved drug related crimes, log total number of crimes in the district reported to the police, and share of those crimes that were violent; (iv) other district characteristics—log population, log area size, log per capita income, log employment, share of people in the district who are religious adherents, female, younger than 20, older than 65, black, white, Hispanic, urban, share with high school education, share with more than high school education, turnout in the most recent presidential election, and total newspaper penetration; and (v) the interaction between all of these variables and judicial selection systems.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 4—REGRESSION OF PENAL PREFERENCES (*Harsh Vote Share*) ON *Congruence*

	All	Nonpartisan	Appointed	Partisan
	(1)	(2)	(3)	(4)
Congruence	−0.007	0.003	0.004	−0.011
	(0.007)	(0.011)	(0.018)	(0.009)
Observations	1,177	556	131	490
R^2	0.879	0.839	0.955	0.881

Notes: This table shows OLS regression results. Standard errors are in parentheses. All columns include state fixed effects, log population, log area size, log per capita income, log employment, crime rate (log number of crimes known to police and the share of violent crimes known to police), education levels (share with high school education and share with more than high school education), share of people in the district who are religious adherents, share black, share urban, turnout in the most recent presidential election, Democratic vote share in the presidential election, and total newspaper penetration. Column 1 also includes dummy variables for appointed and nonpartisan elected judges.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE 5—REGRESSION OF THE SHARE OF CONVICTIONS BY OFFENSE CATEGORY ON *Congruence*

	Homicide, sexual assault, robbery (1)	Homicide, sexual assault, robbery (2)	Violent crimes (3)	Violent crimes (4)	Property crimes (5)	Property crimes (6)	Drug crimes (7)	Drug crimes (8)
Congruence	0.000	0.015	−0.003	0.023	0.012	0.012	0.005	−0.001
	(0.007)	(0.011)	(0.011)	(0.014)	(0.014)	(0.019)	(0.017)	(0.021)
Nonpartisan elected		−0.030**		−0.041*		−0.017		0.068***
		(0.012)		(0.023)		(0.022)		(0.025)
Partisan elected		0.005		−0.000		−0.040		0.037
		(0.017)		(0.022)		(0.032)		(0.038)
Observations	2,860	2,840	2,860	2,840	2,860	2,840	2,860	2,840
R^2	0.355	0.398	0.318	0.351	0.439	0.477	0.465	0.522
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table shows OLS regression results. The dependent variable is the estimated share of convictions in each offense category. The omitted category is weapons and other. The unit of observation is judicial district. Standard errors are in parentheses. State-by-year fixed effects are included in all regressions. Control variables are log population, log area size, log employment, log per capita income, share of religious adherents, female, aged below 20, aged above 65, black, white, Hispanic, urban, schooling, turnout in the most recent presidential election, and total newspaper penetration.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Finally, we investigate whether *Congruence* is related to the underlying distribution of crimes. Table 5 shows the result of regressing the estimated share of convictions in each crime category on *Congruence*. *Congruence* is not significantly correlated with the distribution of the types of crime committed. Thus, the underlying distribution of crimes is unlikely to cause a spurious relationship between *Congruence* and sentencing.[35]

It is useful to note that the absence of a significant relationship between the distribution of crimes and *Congruence* also helps us to rule out the possibility that the relationship between sentencing and *Congruence* might be generated by the behavior of other public officials, e.g., police or prosecutors, altering the composition of criminal cases that are presented before the judge. For example, county sheriffs and prosecutors who face more newspaper monitoring could exert more effort. This could lead to more dangerous criminals being arrested and convicted in comparison with those in districts with less monitoring. In this case, *Congruence* may co-vary with sentencing even absent a response in judicial behavior. The result in Table 5 shows that this theoretical possibility is quantitatively unimportant.

IV. Main Results

In this section, we first analyze how our *Congruence* measure is related to the observed newspaper coverage, as well as how important people say that newspapers

[35] We ran the same regressions with log *Number of Articles* in place of *Congruence* and obtained similar results (no relationships).

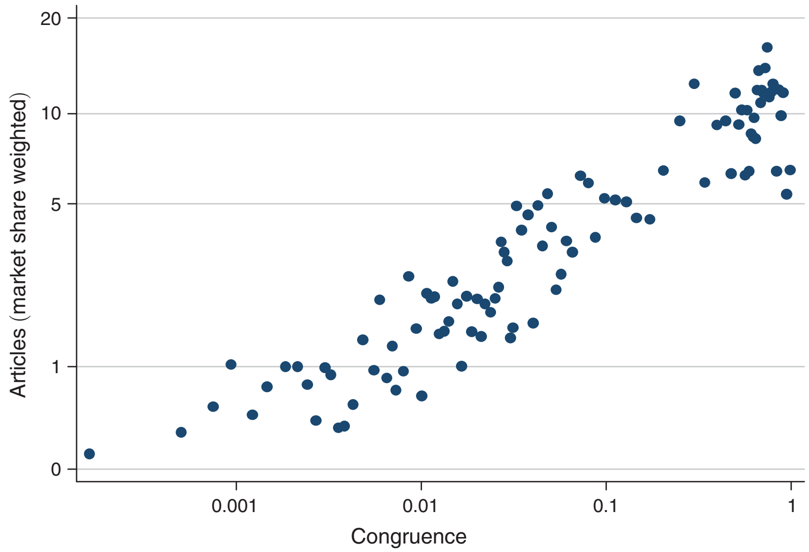

Figure 4. Newspaper Articles and *Congruence*

are for their information about courts. We then turn to the main analysis of the effects of newspaper coverage on sentencing.

A. *Newspaper Coverage*, *Congruence*, *and Voter Information*

We begin by investigating the relationship between *Congruence* and newspaper coverage of the courts. There is a strong positive relationship between *Congruence* and the log *Number of Articles*. This is shown in Figure 4, which displays the binned averages of these two variables. Each dot contains 0.5 percent of the observations, sorted by *Congruence* so that the leftmost dot contains the observations with the lowest 0.5 percent of the observations.

We next investigate this relationship more closely. Column 1 of Table 6 shows the results from a set of regressions of the number of articles per judge per year in a judicial district on *Congruence*. An increase in *Congruence* from zero to one is associated with an additional 23 newspaper articles per judge in the judicial district. This relationship is highly statistically significant. The next column adds a set of control variables: state fixed effects, crime rate, a fixed effect for the Vanderburgh court in Indiana,[36] log population, log per capita income, education (share with 12 years and share with more than 12 years), share black, share urban, log area size, log employment, turnout in the most recent presidential election, Democratic vote share in presidential elections, and share of religious adherents. Among these, share of crimes reported to police that are violent, log population, log per capita income,

[36] Indiana is the only state in this sample that has both partisan and nonpartisan elected judges, and the only court with nonpartisan elected judges in Indiana is Vanderburgh. Vanderburgh is an outlier in terms of news coverage with almost 100 articles, 3 times more than any other court in Indiana. If included, the coefficient of nonpartisan elected judges jumps substantially.

Table 6—Relationship between the Amount of Coverage and *Congruence*

Dependent variable: Articles per judge (market share weighted)	Articles (1)	Articles (2)	log articles (3)
Congruence	22.829***	23.637***	2.195***
	(1.817)	(2.258)	(0.199)
Appointed	−4.383**	−6.272	1.329
	(1.846)	(15.275)	(1.315)
Partisan elected	−0.144	−1.373	1.721
	(1.149)	(16.226)	(1.405)
Observations	1,413	1,413	1,169
R^2	0.102	0.184	0.280
Fixed effects	No	State	State
Controls	No	Yes	Yes

Notes: This table shows OLS regression results. The dependent variable is the market share (circulation) weighted average number of newspaper stories per judge in the court. The unit of observation is judicial district. Standard errors are in parentheses. Columns 2 and 3 include the following control variables: state fixed effects, crime rate (log number of crimes known to police and the share of violent crimes known to police), a fixed effect for the Vanderburgh court in Indiana, log population, log area size, log per capita income, education (share with 12 years and share with more than 12 years), share black, share urban, log employment, turnout in the most recent presidential election, Democratic vote share in the presidential election, and share religious adherents. The number of observations is reduced in column 3 because of observations with zero articles and the logarithm.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

and share of religious adherents are positively related to number of newspaper stories. The estimated relationship between *Congruence* and number of articles is not affected much by these controls. In sum, our estimates suggest that an increase in *Congruence* from 0 to 1 is associated with an additional 24 newspaper articles per judge in the judicial district after adding controls. A 1 standard deviation increase in *Congruence* implies effects about one-third as large, or around eight more articles.

We next investigate how voter information about courts is related to our measures of newspaper coverage. The National Center for State Courts conducted a survey in 2000 where they asked a random sample of US respondents the following question: "How important to you are the following sources of information to your overall impression of how the courts in your community work?" They were given ten alternatives including newspapers, television news, their own experience in court, friends, relatives, their job, etc. We regressed a dummy variable for whether the respondent said that newspapers were very or somewhat important on the log *Number of Articles* and on *Congruence*. The results are shown in Table 7. Respondents in judicial districts where *Congruence* is high are about 35 percentage points more likely to cite newspapers as an important source of information about the courts, and this is highly statistically significant and robust to the inclusion of controls. The implied effects per newspaper story are large, as an additional 24 newspaper articles are associated with a 35 percentage point increase in the share of respondents who report newspapers being an important source of information. This is consistent with the down-ballot nature of judicial elections, with voters who are largely uninformed and influenced by limited media coverage.

Table 7—Relationship between Voter Information, *Congruence*, and Amount of Coverage

Dependent variable: Newspapers important for information about courts	(1)	(2)	(3)	(4)	(5)
Congruence	0.345***	0.394***			0.436***
	(0.095)	(0.103)			(0.130)
log *number of articles*			0.021*	0.024	0.026*
			(0.012)	(0.015)	(0.015)
Observations	533	531	475	473	473
R^2	0.123	0.155	0.115	0.146	0.172
Controls	No	Yes	No	Yes	Yes

Notes: This table shows OLS regression results. The unit of observation is survey response. All specifications include state-fixed effects. Standard errors, clustered by judicial district, are in parentheses. The control variables are log population, log area size, log per capita income, share of population younger than 25, older than 65, urban, black, white, Hispanic, share with high school education, share with more than high school education, and turnout in the most recent presidential election. The number of observations is reduced in columns 3–5 because of observations with zero articles and the logarithm.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Respondents in judicial districts with many newspaper articles covering judges are also more likely to cite newspapers as an important source of information. However, this effect is only marginally statistically significant. One reason for the weaker result could be that the survey was conducted in 2000, while the data we collected on the number of newspaper articles covering judges was for 2004 and 2005.

In sum, we have found that *Congruence* is strongly and positively correlated with the number of newspaper articles written about judges and with people stating that newspapers are an important source of information about the courts. We now turn to the effects on sentencing.

B. *Effects of Newspaper Coverage on Sentencing*

We now investigate how newspaper coverage influences sentencing. In Section IVB, we focus on the three most serious offense types—homicides, sexual assaults, and robberies—because these are most likely to attract media attention. We first look at the average effects of newspaper coverage across all judicial selection systems. Then, in Section IVB, we analyze the effects of newspaper coverage by judicial selection system. Finally, in Section IVB, we analyze the effect for less severe crimes.

All Selection Systems—Most Severe Crimes.—We first restrict attention to homicides, sexual assaults, and robberies. Panel A of Table 8 presents estimates from regressions of *Harshness* on our media variables for these crimes. All specifications include state-by-year fixed effects. The first three columns use the log *Number of Articles* about judges in 2004 and 2005 as the main independent variable.[37] The first

[37] Since we have the number of articles only for 2004 and 2005 but use sentencing data from 1986 to 2006, we use the number of articles per judge per year in 2004 and 2005 for the observations from other years. Therefore, the results from the first three columns yield only a rough, overall relationship between newspaper coverage and sentencing, rather than a precise relationship. Moreover, we are mainly interested in the relationship between

TABLE 8—REGRESSION OF *Harshness* ON MEDIA VARIABLES, PENAL PREFERENCES, AND SELECTION SYSTEMS

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Overall relationships between media coverage and sentencing						
log number of articles	0.026***	0.017***	0.015***			
	(0.004)	(0.004)	(0.004)			
Congruence				0.027	0.051**	0.061**
				(0.019)	(0.023)	(0.025)
Harsh vote share		0.113	0.105		0.142	0.179*
		(0.090)	(0.102)		(0.091)	(0.099)
Democratic vote share		−0.013	−0.017		−0.021	0.013
		(0.044)	(0.051)		(0.042)	(0.041)
Appointed		−0.086***	−0.110***		−0.102***	−0.089**
		(0.032)	(0.037)		(0.023)	(0.035)
Partisan elected		−0.011	−0.022		−0.022	
		(0.025)	(0.015)		(0.027)	
Observations	212,837	147,497	142,363	232,470	163,551	100,983
R^2	0.133	0.131	0.131	0.128	0.128	0.115
Controls	No	Yes	Yes	No	Yes	Yes
Trimmed sample	No	No	Yes	No	No	Yes
Indirect least squares				0.034	0.080	0.097

	Nonpartisan		Appointment		Partisan	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B. Media influence by judicial selection system						
Congruence	0.022	0.098***	−0.024	−0.022	0.029	−0.004
	(0.038)	(0.036)	(0.032)	(0.028)	(0.021)	(0.027)
Harsh vote share		0.272*		−0.685***		0.069
		(0.151)		(0.240)		(0.134)
Democratic vote share		−0.074		−0.011		−0.056
		(0.076)		(0.082)		(0.065)
Observations	109,414	95,515	29,234	24,120	93,822	43,916
R^2	0.114	0.126	0.068	0.074	0.156	0.146
Controls	No	Yes	No	Yes	No	Yes
Trimmed sample	No	No	No	No	No	No
Indirect least squares	0.028	0.155	−0.030	−0.035	0.037	−0.007

Notes: This table shows OLS regression results. The unit of observation is an individual felony case. Standard errors, clustered by judicial district, are in parentheses. All specifications include state-by-year fixed effects. The number of observations is smaller in columns with control variables because *Harsh Vote Share*, based on ballot propositions, is not available in all states. Control variables are (i) fixed effects for type of crime; (ii) defendant characteristics—gender, race, age, and age squared; (iii) crime-related district characteristics—log total number of convictions in the district, share of convictions that involved violent crimes, and share that involved drug related crimes, log total number of crimes in the district reported to the police, and share of those crimes that were violent; and (iv) other district characteristics—log population, log area size, log per capita income in the district, log employment, the share of people in the district who are religious adherents, female, younger than 20, older than 65, black, white, Hispanic, and urban, share with high school education, share with more than high school education, turnout in the most recent presidential election, and total newspaper penetration. In column 6 in panel A, partisan election dummy is omitted because of multicollinearity. Tables A.10 and A.12 in Section F of the online Appendix present results from various sensitivity analyses of columns 2 and 5 in panel A. Tables A.11 and A.13 in the same section present results in panel B using interactions between selection systems and *Congruence*, as well as their sensitivity analyses. The bottom row of each panel, labeled "Indirect Least Squares," shows the *Congruence* coefficient scaled by the first stage effect of *Congruence* on log *Number of Articles*.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

sentencing and *Congruence*, which is less subject to endogeneity. Thus, results from the first three columns are useful primarily for the purpose of comparison with results from using *Congruence* as the key independent variable.

column shows that the number of newspaper articles is positively and significantly associated with harsher sentencing. This correlation could exist for a variety of reasons. One is that more severe crimes attract media attention. However, the correlation does not seem to be driven by specific cases during the time period of our newspaper data. The correlation is virtually unchanged when we remove 30,000 observations from the year 2004 (coefficient estimate = 0.0255, s.e. = 0.0044).

Another possible reason for the correlation between sentencing and newspaper coverage is that newspapers are likely to be located in densely populated areas, and these areas might also have different crime rates and sentencing patterns. Areas with ample newspaper coverage may also differ in demographic characteristics such as race, ethnicity, income, and education. To address such possibilities, in column 2 we include a large set of control variables, identical to those used in Section IIIB: (i) fixed effects for type of crime; (ii) defendant characteristics—gender, race, age, and age squared; (iii) crime-related district characteristics—log total number of convictions in the district, share of convictions that involved violent crimes, share that involved drug related crimes, log total number of crimes in the district reported to the police, and share of those crimes that were violent; and (iv) other district characteristics—log per capita income, log employment, share of people in the district who are religious adherents, female, younger than 20, older than 65, black, white, Hispanic, urban, share with high school education, share with more than high school education, turnout in the most recent presidential election, and total newspaper penetration. The inclusion of these controls reduces the estimated coefficient of the log *Number of Articles* substantially compared to column 1.[38]

The third specification, in column 3, uses a trimmed sample. This excludes all observations which, based on population and area size, have a less than 10 percent or above 90 percent probability of having the log *Number of Articles* above median. The estimated coefficient of the log *Number of Articles* falls even further.

The next three columns repeat the same specifications, but use *Congruence* as the key independent variable. When we include a large set of controls as above in column 5, the estimated coefficient of *Congruence* is large and statistically significant.[39] The final specification, in column 6, uses a trimmed sample excluding all observations which, based on population and area size, have a less than 10 percent or above 90 percent probability of having *Congruence* above median. The estimated coefficient is of similar size and significant in this trimmed sample.

Given the results presented in column 3 of Table 6, we would expect the coefficient of *Congruence* to be about 2.2 times as large as the coefficient of the log *Number of Articles*. The actual coefficient estimates of *Congruence* are larger, but not significantly so.[40]

[38] We estimated a specification controlling for only two variables—log population and log area size. Including these two variables alone causes the estimated coefficient to fall substantially. We obtained 0.020 for the coefficient estimate of log *Number of Articles* with standard error 0.003.

[39] When we control for only log population and log area size, we obtain 0.046 for the coefficient of *Congruence* and 0.022 for its standard error.

[40] We do not have a strong prior on the direction of bias with the log *Number of Articles*. On one hand, areas with more severe crimes may get more articles about judges and longer sentences producing a positive bias. On the other hand, we only measure newspaper coverage for two years covering 13 percent of the observations in our sample. So there is most likely considerable measurement error and consequent attenuation bias.

Several empirical papers in the literature argue that elected judges are generally harsher than appointed judges. For example, Gordon and Huber (2007) compare elected and appointed judges in Kansas and argue that the former impose longer sentences. We find similar differences. In Table 8, appointed judges are associated with less harsh sentencing. There is no discernible difference between partisan and nonpartisan elected judges.[41]

We also estimated a specification that includes an interaction term between *Congruence* and *Harsh Vote Share*, but this term was not statistically significant. Theoretically, the effect of *Congruence* on *Harshness* may not necessarily increase in *Harsh Vote Share* because localities that have harsher penal preferences will select judges with such preferences in the first place. That is, judges' intrinsic preferences already reflect voters' preferences to some extent even without media influence. Thus, the marginal change in sentencing behavior that media coverage causes may not vary much across the preference of localities.

Regarding the controls, male and black convicts receive significantly harsher sentences. There is a strong age profile where convicts at age 44 get the longest sentences. *Harshness* is not correlated with the number or the types of crimes dealt with in the courts or reported to the police, given that we compare sentencing only within the group of cases that have common penal code citations. *Harshness* is higher in districts that are rich and with small area size.

Effects by Judicial Selection System.—We now investigate whether the influence of newspaper coverage depends on the judicial selection system. We argued in Section IB that newspapers would have larger effects on sentencing for nonpartisan elected judges, compared to partisan elected and appointed judges. To study these differences, we regressed *Harshness* on *Congruence* and a large set of control variables specified above separately for each judicial selection system. As in the previous subsection, we focus on severe crimes—homicides, sexual assaults, and robberies.

The results from these regressions are shown in panel B of Table 8. Columns 1 and 2 show the results for the nonpartisan system without and with control variables, respectively. Column 2 shows a large and statistically significant effect of *Congruence* on *Harshness*. The effect is about twice as large as the average effect across selection systems that we measured in panel A (0.098 compared to 0.051). Consistent with our argument, the effect is larger for nonpartisan elected judges than for partisan elected or appointed judges. Using the estimate of 0.098, a 1 standard deviation increase in *Congruence* is estimated to increase the sentence length by about 5.7 months (3.4 percent) in the nonpartisan election system.[42] In the bottom row of each panel, labeled "Indirect Least Squares," we present indirect least squares coefficients, that is, the *Congruence* coefficients scaled (divided) by the

[41] We can identify the coefficients of partisan elected and appointed judges because some states have within-state variation in judicial selection system.

[42] To relate sentence length to *Harshness*, we regress the former on the latter yielding a coefficient of 194 (controlling for state-by-year fixed effects and offense category). This is multiplied by the estimated effect of *Congruence* on *Harshness* (0.098), and 1 standard deviation of *Congruence* (0.30), yielding $194 \times 0.098 \times 0.3 = 5.7$ months.

first stage effect of *Congruence* on the log *Number of Articles* obtained from the data used in Table 8. The increase in *Harshness* in response to a 1 standard deviation increase in log *Number of Articles* for nonpartisan elected judges, based on the indirect least squares coefficient, is 21.3 months (12.8 percent), which is even larger.[43] In contrast, we find no evidence on the influence of newspaper coverage in the appointment (columns 3 and 4) or the partisan election (columns 5 and 6) systems. The estimated effect of *Congruence* is 0.120 and 0.102 points lower for appointed and partisan elected judges, respectively, than for nonpartisan elected judges.

In Table A.11 of the online Appendix, we present these results using interactions between *Congruence* and judicial selection systems instead of sample splits. There we test statistical significance of the differences between systems. The differences between the nonpartisan and the other systems are statistically significant. The difference between the appointment and the partisan election systems are not statistically significant.

We interpret the differences in the estimated media influence between the nonpartisan election and the other systems to be caused by differential selection and incentive effects. A priori, there are other possibilities. Hanssen (2004a) argues that politically unstable states are more likely to appoint judges. In those states, there would likely be more swing voters inducing larger media effects. Consequently, the difference in the estimated coefficient of *Congruence* between the nonpartisan election and the appointment systems would capture the difference in voter characteristics as well as media influence. However, we do not find a positive media influence on appointed judges, which mitigates this concern. Another possibility is that judicial districts with nonpartisan elected judges have more severe and, hence, newsworthy crimes, or perhaps they have more Republican newspapers, which are more anti-defendant or cover the courts more. However, we find little evidence of these. Table 5 shows that, if anything, the share of violent crimes is lower in nonpartisan elected districts. In Section G of the online Appendix, we show that newspaper partisanship is not correlated with the judicial selection system.

The zero result for appointed and partisan elected judges cannot be generalized to high-profile offices with significantly more voluminous media coverage. However, they do indicate an interesting limit to media effects in low-information offices.

We interpret the positive media influence on sentencing harshness as evidence that those informed by the media prefer harsher sentences, and that the media help them enforce this preference.[44] The large effects may arise because voters are largely uninformed about judicial elections, so even one publicized case can decide their votes, and because the media acts as a "fire alarm" (see, e.g., McCubbins and Schwartz 1984), that is, the potential threat of negative coverage keeps judges in check.

[43] This calculation is analogous to that in footnote 42: $194 \times 0.155 \times 0.71$ (from panel B of Table 2) $= 21.3$ months. This is much larger than the 5.7 month figure mentioned above, because the change in the log *Number of Articles* in response to a 1 standard deviation increase in *Congruence* is less than $1/3$ standard deviation of the log *Number of Articles*.

[44] If newspaper owners and editors themselves prefer harsh sentences, and newspaper endorsements of judges have a large effect only in nonpartisan elections, it would also explain our results. However, since we have systematic survey data on voters' penal preferences, while we have no data on penal preferences of newspaper owners and editors, we attribute our results to the former.

Table 9—Heterogeneous Effects by Offense Category and Race of Defendants for Nonpartisan Elected Judges

	Homicide, sexual assault, robbery (1)	Violent (2)	Property (3)	Drug (4)	Weapon and other (5)
Panel A. By offense category only					
Congruence	0.098***	0.076***	0.069***	0.052*	0.039*
	(0.036)	(0.029)	(0.026)	(0.029)	(0.023)
Observations	95,515	232,227	354,813	452,226	215,820
R^2	0.126	0.093	0.057	0.086	0.050
Indirect least squares	0.155	0.121	0.109	0.082	0.062
Panel B. By offense category and race of defendants					
Congruence	0.103***	0.079***	0.071**	0.045	0.040*
	(0.037)	(0.030)	(0.028)	(0.029)	(0.024)
Congruence × black defendant	−0.017	−0.009	−0.008	0.024	−0.006
	(0.023)	(0.019)	(0.022)	(0.029)	(0.020)
Observations	95,515	232,227	354,813	452,226	215,820
R^2	0.126	0.093	0.057	0.086	0.050
Indirect least squares	0.163	0.124	0.111	0.071	0.064

Notes: This table shows OLS regression results. The unit of observation is an individual felony case. Standard errors, clustered by judicial district, are in parentheses. All specifications include state-by-year fixed effects and control variables. The set of control variables is the same as that in Table 8. The number of observations is smaller in columns 1–5 because of observations with zero articles and the logarithm. Violent crimes include murder, sexual assault, robbery, aggravated assault, and other violent crimes. Property crimes include burglary, larceny and fraud. Drug crimes include drug possession and drug trafficking. Weapons include weapon offenses and other offenses. The bottom row of each panel, labeled "Indirect least squares," shows the *Congruence* coefficient scaled by the first stage effect of *Congruence* on log *Number of Articles*.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Robustness Checks.—In Section F of the online Appendix, we document a number of robustness checks for the results presented in Table 8. Ohio has a unique system with partisan primaries and nonpartisan general elections, and Maryland has a mix of a nonpartisan and appointment system. Our results are not sensitive to coding Ohio as partisan rather than nonpartisan or Maryland as appointment rather than nonpartisan. Texas is the only state where the jury can decide the final sentence upon defendants' requests.[45] Our results are not sensitive to dropping Texas from the analysis.[46]

We also performed additional checks to deal with the mechanical correlation of *Congruence* with city size. We excluded all districts that are 100 percent urban

[45] In practice, most criminal cases are resolved through plea bargaining. Thus, the proportion of cases in which the jury decides the final sentence is negligible.

[46] We also checked robustness of the results to excluding California, which is regarded as particularly stringent in sentencing policies. When we exclude California, the coefficient estimate for *Congruence* loses statistical significance and the magnitude becomes smaller, although we still get point estimates that show larger effects under the nonpartisan system. However, stringent sentencing policies in California are unlikely to be a confounding factor. Our measure of sentencing harshness is normalized within state. Thus, cross-state heterogeneity of the overall variation in sentencing harshness is removed to a large extent in the normalization process. Moreover, we include state-year fixed effects in all our analysis of sentencing decisions, which implies that our main result is based on comparison of judicial districts within state.

(around 7 percent of the observations). We added controls constructed similar to *Congruence*, but replaced $ReaderShare_{md}$ with the population of newspapers' headquarter market weighted by circulation. We finally controlled for demographics at the MSA-level. These did not significantly change our results.

Partisanship of Newspapers.—One remaining concern in interpreting our main results is the possibility that the influence of newspaper coverage differs across judicial selection systems because newspaper ideologies do. We address this issue in Section G of the online Appendix. We measure partisanship of newspapers by counting the number of endorsements from each newspaper for Democratic and Republican candidates in presidential elections. We find that newspaper partisanship is not correlated with judicial selection systems. Moreover, we find no systematic difference in the amount of newspaper coverage about courts or the magnitude of the effect between Democrat-leaning and Republican-leaning newspapers.

Effects by Severity.—Some types of criminal cases attract more media attention than others. For example, in one content analysis of news reports, murder accounted for 25 percent of crime stories, although it constituted less than 1 percent of all reported crimes (Graber 1988). Similarly, in a search on NewsLibrary.com, we identified more than 2 million newspaper articles mentioning "judge" and "sentenc*." Forty-two percent of the newspaper articles mention the most severe violent crimes—homicides, sexual assaults, and robberies—although they constituted only 7 percent of all felony convictions in 2006 (Rosenmerkel, Durose, and Farole 2009). Consequently, the newspaper coverage to convictions ratio for these crimes is $42/7 = 6$. This ratio is 2.89 for all violent crimes, 0.86 for property crimes (burglary, theft, fraud), and 0.76 for drug related crimes.[47] This implies that newspaper coverage should have the greatest effect on the most severe violent crimes. If ordinary voters prefer longer sentences, *Congruence* should correlate more with harsh sentencing for these cases.

To investigate this, we run separate regressions by type of crime. The results are shown in panel A of Table 9. The specifications include controls and the full (nontrimmed) sample, similar to column 5 in panel A of Table 8.

The estimated coefficients of *Congruence* are higher for more newsworthy offense categories with higher ratios of newspaper coverage to convictions. The estimated effect is statistically significant for violent crimes and property crimes, but smaller than for the most severe violent crimes in the first column. The correlation between *Congruence* and sentencing harshness is driven by violent crimes. The size of the coefficient drops sharply and becomes insignificant at the 5 percent significance level for drug or weapon-related crimes.

There is also some evidence that black defendants are overrepresented in media coverage. For example, Dixon and Linz (2000) find that while 25 percent of all

[47] Violent crimes are mentioned in 52 percent of the articles and constitute 18 percent of convictions. Property crimes (burglary, theft, or fraud) are mentioned in 24 percent of the articles and constitute 28 percent of the convictions. Drugs are mentioned in 25 percent of the articles and drug crimes constitute 33 percent of the convictions.

felony perpetrators according to crime reports were black, 44 percent of the perpetrators on television news were black. By comparison, 23 percent of the felony perpetrators were white, while only 18 percent of felony perpetrators on television news were white. Again, if effects on sentencing are increasing in media coverage, we would expect to see a stronger correlation between *Congruence* and sentencing in cases involving black defendants. However, we found no evidence of this: see panel B of Table 9.

To get a sense of the aggregate implications of changes in the media environment, consider a uniform increase in *Congruence* by 1 standard deviation affecting all the 108,535 estimated convictions of violent crimes and 170,758 convictions of property crimes under nonpartisan elected judges in 2006. The estimates imply that this change would have increased the aggregate sentence length for these crimes by more than 50,000 years.[48]

V. Conclusion

This paper studies interaction between media environments and political institutions that govern selection and incentives of public officials. It provides empirical evidence on how newspaper coverage influences the behavior of US state court judges, for whom selection systems vary across states. The core result shows that newspaper coverage of courts significantly increases sentence length by nonpartisan elected judges for severe crimes.

This study differs significantly from previous studies on media influence in two dimensions: (i) we highlight the role of institutional design, electoral systems in particular, in media influence on the behavior of public officials; and (ii) we analyze media influence on low-information, down-ballot offices. The amount of newspaper coverage on down-ballot offices differs from that on major politicians. The number of newspaper articles that we find covering judges is only roughly 10 percent and 1 percent of those found covering congressmen and governors, respectively. Not surprisingly, many voters report not having enough information about judicial candidates. In this type of setting, voters have weak priors about candidates. Thus, even a small increment in newspaper coverage may have a significant influence. At the same time, the influence of media coverage can also be crowded out easily by a different type of information provided to voters, such as candidates' party affiliation on the ballot in the partisan election system. Similarly, when voters can have only indirect influence on the selection of judges, as in the appointment system, newspaper coverage would not matter because of the importance of nonjudicial issues in the gubernatorial elections.

Our findings have broad implications because many local officials are selected in a manner analogous to the process used for selecting judges. For example, public utility regulators and school board superintendents also have similar variation in

[48] For each crime category, we regress sentence length on *Harshness*, controlling for state-by-year fixed effects and offense category. The coefficient of *Harshness* is multiplied by the estimated effect of *Congruence* on *Harshness* from Table 9, 1 standard deviation of *Congruence* (0.30), and the estimated number of convictions in the offense category under nonpartisan elected judges.

their selection rules across states. Our result shows that selection systems interact significantly with media environments in influencing the behavior of public officials in low-information offices.

Future research could go in two directions. One direction would be to study the diffusion of other types of media. For example, the influence of Internet and social media on low-information offices may be very different from that of newspaper coverage. On one hand, diffusion of Internet and social media may have a greater influence on low-information offices because they make information flow faster. On the other hand, those new media are less likely to cover low-information offices because of their focus on entertainment-oriented contents. The other direction would be to study other low-information offices for which the nature of tasks differs from that of the judiciary. For example, state legislators get relatively little media coverage compared to national politicians. Their task also differs significantly from the state judges we study, in that legislators make policies of a highly partisan nature or directly engage in resource allocation across constituents with conflicting interests. In such a setting, the influence of media coverage may well be different from that on the state judges we study or the national politicians that previous studies have focused on. Research on these issues would enrich our understanding of the interaction between media and political environments.

REFERENCES

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113 (1): 151–84.

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab. 1995. "Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes." *Journal of Legal Studies* 24 (2): 257–81.

Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14 (1): 19–42.

Besley, Timothy, and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117 (4): 1415–51.

Besley, Timothy, and Stephen Coate. 2003. "Elected Versus Appointed Regulators: Theory and Evidence." *Journal of the European Economic Association* 1 (5): 1176–1206.

Bonneau, Chris W., and Melinda Gann Hall. 2009. *In Defense of Judicial Elections.* New York: Routledge.

Brace, Paul, and Melinda Gann Hall. 1997. "The Interplay of Preferences, Case Facts, Contexts, and Rules in the Politics of Judicial Choice." *Journal of Politics* 59 (4): 1206–31.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122 (3): 1187–1234.

Dixon, Travis L., and Daniel Linz. 2000. "Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News." *Journal of Communication* 50 (2): 131–54.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Schleifer. 2003. "Courts." *Quarterly Journal of Economics* 118 (2): 453–517.

Dyck, Alexander, David Moss, and Luigi Zingales. 2013. "Media versus Special Interests." *Journal of Law and Economics* 56 (3): 521–53.

Eisensee, Thomas, and David Strömberg. 2007. "New Floods, News Droughts, and U.S. Disaster Relief." *Quarterly Journal of Economics* 122 (2): 693–728.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review* 101 (7): 3253–85.

Ferejohn, John. 1986. "Incumbent performance and electoral control." *Public Choice* 50 (1–3): 5–25.

Ferraz, Claudio, and Frederico Finan. 2011. "Electoral Accountability and Corruption: Evidence from the Audits of Local Governments." *American Economic Review* 101 (4): 1274–1311.

Gentzkow, Matthew. 2006. “Television and Voter Turnout.” *Quarterly Journal of Economics* 121 (3): 931–72.

Glaeser, Edward L., and Andrei Schleifer. 2002. “Legal Origins.” *Quarterly Journal of Economics* 117 (4): 1193–1229.

Goldman, Sheldon. 1966. “Voting Behavior on the United States Courts of Appeals, 1961–1964.” *American Political Science Review* 60 (2): 374–83.

Goldman, Sheldon. 1975. “Voting Behavior on the United States Courts of Appeals Revisited.” *American Political Science Review* 69 (2): 491–506.

Gordon, Sanford C., and Gregory A. Huber. 2007. “The Effect of Electoral Competitiveness on Incumbent Behavior.” *Quarterly Journal of Political Science* 2 (2): 107–38.

Graber, Doris A. 1988. *Processing the News: How People Tame the Information Tide.* New York: Longman Group.

Hall, Melinda Gann. 2001. “State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform.” *American Political Science Review* 95 (2): 315–30.

Hanssen, F. Andrew. 1999. “The Effect of Judicial Institutions on Uncertainty and the Rate of Litigation: The Election versus Appointment of State Judges.” *Journal of Legal Studies* 28 (1): 205–32.

Hanssen, F. Andrew. 2000. “Independent Courts and Administrative Agencies: An Empirical Analysis of the States.” *Journal of Law, Economics, and Organization* 16 (2): 534–71.

Hanssen, F. Andrew. 2004a. “Is There a Politically Optimal Level of Judicial Independence?” *American Economic Review* 94 (3): 712–29.

Hanssen, F. Andrew. 2004b. “Learning about Judicial Independence: Institutional Change in State Courts.” *Journal of Legal Studies* 33: 431–73.

Huber, Gregory A., and Sanford C. Gordon. 2004. “Accountability and Coercion: Is Justice Blind when It Runs for Office?” *American Journal of Political Science* 48 (2): 247–63.

Lim, Claire S. H. 2013. “Preferences and Incentives of Appointed and Elected Public Officials: Evidence from State Trial Court Judges.” *American Economic Review* 103 (4): 1360–97.

Lim, Claire S. H., Bernardo Silveira, and James M. Snyder, Jr. 2015. “Do Judges’ Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts.” https://lim.economics.cornell.edu/texaspaper.pdf.

Lim, Claire S. H., and James M. Snyder, Jr. 2015. “Is More Information Always Better? Party Cues and Candidate Quality in U.S. Judicial Elections.” *Journal of Public Economics* 128: 107–23.

Lim, Claire S. H., James M. Snyder, Jr., and David Strömberg. 2015. “The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems: Dataset.” *American Economic Journal: Applied Economics*. http://dx.doi.org/10.1257/app.20140111.

McCubbins, Mathew D., and Thomas Schwartz. 1984. “Congressional Oversight Overlooked: Police Patrols versus Fire Alarms.” *American Journal of Political Science* 28 (1): 165–79.

Nagel, Stuart. 1961. “Political Party Affiliation and Judges’ Decisions.” *American Political Science Review* 55: 843–50.

Pinello, Daniel R. 1999. “Linking Party to Judicial Ideology in American Courts: A Meta-analysis.” *Justice System Journal* 20 (3): 219–54.

Prat, Andrea, and David Strömberg. 2013. “The Political Economy of Mass Media.” In *Advances in Economics and Econometrics: Theory and Applications, Proceedings of the Tenth World Congress of the Econometric Society*, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Cambridge: Cambridge University Press.

Rosenmerkel, Sean, Matthew Durose, and Donald Farole, Jr. 2009. *Felony Sentences in State Courts, 2006: Statistical Tables*. Bureau of Justice Statistics. Washington, DC, December.

Sheldon, Charles H., and Nicholas P. Lovrich, Jr. 1999. “Voter knowledge, Behavior and Attitudes in Primary and General Judicial Election.” *Judicature* 82 (5): 216–23.

Snyder, James M., Jr., and David Strömberg. 2010. “Press Coverage and Political Accountability.” *Journal of Political Economy* 118 (2): 355–408.

Streb, Matthew J., Brian Frederick, and Casey LaFrance. 2007. “Contestation, Competition, and the Potential for Accountability in Intermediate Appellate Court Elections.” *Judicature* 91 (2): 70–78.

Strömberg, David. 2004. “Radio’s Impact on Public Spending.” *Quarterly Journal of Economics* 119 (1): 189–221.

Tabarrok, Alexander, and Eric Helland. 1999. “Court Politics: The Political Economy of Tort Awards.” *Journal of Law and Economics* 42 (1): 157–88.

Tate, C. Neal. 1981. “Personal Attribute Models of Voting Behavior of U.S. Supreme Court Justices: Liberalism in Civil Liberties and Economics Decisions, 1946–1978.” *American Political Science Review* 75 (2): 355–67.