Of Challengers and Constituents: The Effects of Institutional Design on Judicial Congruence with Public Opinion

Michael J. Nelson*
Department of Political Science
The Pennsylvania State University

April 9, 2015

Abstract

Perhaps the most important question in the study of judicial politics in the American states concerns the extent to which judicial responsiveness to public opinion is mitigated and exacerbated by institutional arrangements. Past attempts to answer this question have either relied upon cross-sectional comparisons—which leave unaddressed the possibility that what appears to be congruence with public opinion is actually a function of unmeasured legal or cultural considerations—drawn from collegial courts, where interpersonal dynamics and nonindependence among justices' votes may bias estimates of congruence. In this paper, I address these concerns by adapting and validating a constituency-level measure of public attitudes toward punitiveness and investigating the congruence of trial court judges with public opinion in a state in which judges are both elected and appointed. The results indicate that, among typical levels of punitiveness, judges elected in partisan elections sentence more severely than those retained via uncontestable elections.

*This paper was prepared for the 2015 meeting of the Midwest Political Science Association. This paper is a draft; please do not cite without permission.

Introduction

It is a truism that the American people prefer judges who are tough on crime. As former Oregon Supreme Court justice Hans Linde (1987) has written "Every judge's campaign slogan, in advertisements and on billboards, is some variation of 'tough on crime"' (2000).[1] Many of the highest-profile anti-retention campaigns, such as the successful ones against Tennessee's Penny White and California's Rose Byrd, have emphasized justices' failures to vote to put criminal defendants to death. The anecdotal evidence in this point is bolstered by systematic empirical evidence. Hall (2001) has documented that voters in judicial elections cast their ballots retrospectively on the basis of violent crime.

Yet, Americans' attitudes toward crime are neither unidirectional nor constant. Since the 1970s, the General Social Survey has asked Americans about whether they believe courts are too harsh, not harsh enough, or about right when dealing with criminals. While the majority of Americans, according to the GSS, have always believed that the judicial system is too lenient, the percentage has declined from 90.2% of the public in 1978. Today, about three-in-five Americans feel that the court system needs to be more punitive. The percent of the American people who believe the judiciary is too punitive has increased over fivefold from 1994 to today.[2]

This temporal change in attitudes toward punitiveness mask cross-sectional variation across the country. As Enns (2014) has documented, attitudes toward punitiveness vary cross-sectionally, with some states substantially more punitive than others. While it is undeniable that a sizable portion of the American public favors more punitive sentences for criminals, it is similarly clear that a non-insignificant segment of the American people hold opposing views.

The consequences of these public attitudes on judges' decisions are unclear. Judges in the American states are selected and retained through a variety of appointive and electoral methods, and perhaps the most important question in the study of judicial politics in the American states

[1] As an extreme example, Blume and Eisenberg (1998) report that an "Alabama judge sought reelection with the campaign slogan: "Some criminal complain that he's too tough on criminals, AND HE IS.... We need him now more than ever" (475).

[2] More evidence comes from the 2000 Annenberg National Election Studies, where about 25% of the American people answered that underpunishment is neither "not a problem" nor a "not too serious" problem.

is the extent to which judicial congruence with public opinion is mitigated and exacerbated by institutional arrangements. The bulk of the evidence indicates that judges who are subject to judicial elections are more congruent with public opinion than those who are not (Brace and Boyea 2008) and that the type of judicial election used to retain a judge conditions the level of congruence (Canes-Wrone, Clark and Kelly 2014). Yet, exactly *how* method of judicial retention conditions the effect of public opinion on judicial sentencing is unclear. Canes-Wrone, Clark and Kelly (2014) suggest that public opinion exerts a greater pull upon judges who face nonpartisan elections, while Gordon and Huber (2007) find that the presence of party labels on the ballot is associated with more punitive sentences.

Reconciling these findings is difficult to do because judges on the same court are nearly always retained by the same procedure.[3] Scholars have attempted to side-step this problem by examining the effects of public opinion on judicial decisionmaking cross-sectionally using decisions of state supreme courts. Yet, law both affects judicial decisionmaking (Bailey and Maltzman 2011) and varies across jurisdictional boundaries. Moreover, opinions on appellate courts are influenced by other judges who sit on the panel (Maltzman, Spriggs II and Wahlbeck 2000), leaving open the possibility that appellate judges influence each other to rule in favor of public opinion, thus overstating the influence of public opinion on judicial decisionmaking. State trial courts in Kansas offer a unique solution this problem: about half of these judges face contestable elections before they will be retained in office while the other half face uncontestable retention elections. All of these judges, however, make unilateral decisions and are subject to the same laws, mitigating the difficulties posed by studies of appellate courts.

In this study, I examine the conditional effects of public attitudes toward punitiveness on the sentences of Kansas district court judges. I find that, contrary to studies of appellate courts, but following Gordon and Huber (2007), judges elected in partisan elections sentence more severely than those retained via uncontestable elections, at least among typical levels of punitiveness. The results have important implications for our understanding of the effects of

[3]Mandatory retirement ages provide an important exception, as Hall (2014) has shown.

judicial elections. In particular, they suggest that the consequences of institutional design for the congruence of courts with public opinion varies across the judicial hierarchy.

Crime, Punitiveness and Judicial Elections

The linkage between attitudes toward punitiveness and judicial elections is not a new one. Champagne (2001) argues that the now-usual approach to judicial elections began when deputy district attorneys in southern California began to encourage voters to oppose judges they viewed as "soft on crime" (669). In recent years, Bright and Keenan (1995) argue, "[a]n opponent can seize upon a judge's ruling in one case and, by focusing on the facts of the crime and completely ignoring the legal issue, make even the toughest judge appear 'soft on crime"' (785). Baum (2003) goes so far as to claim that "creating the impression that a judge is soft on crime can have great electoral impact" (35).

Judging on the television advertisements they run today, judges seem to believe that being viewed as punitive will help them on election day. Champagne (2001) quotes a typical early television advertisement for a State Supreme Court race:

> Why did the Alabama Fraternal Order of Police endorse Judge **** over **** for Alabama Supreme Court? Because she respects law enforcement. Judge ****. A twenty-year record fighting crime as a prosecutor and judge. A ninety-one percent conviction rate in DUI cases as a district judge. And last year in two tragic cases, Judge **** sentenced two convicted murders to the death penalty" (677). Indeed, crime is unquestionably the most important issue in judicial campaigns (Hall 2001).

Champagne also quotes another 2000 advertisement from Michigan in which the candidate declared that "I think for too long our courts have really emphasized the rights of criminals at the expense of victims" and the ad's announcer stated that "Supreme Court Justice **** believes in protecting the rights of police officers, victims, and law-abiding citizens. Technicalities or loopholes shouldn't keep criminals on the street. That is why Justice **** is supported by more than 22,000 Michigan police officers" (677).

Huber and Gordon (2004) suggest that judges reflect the public's punitive stance in their sentences as well as their campaign advertisements. Because a single "bad" sentence can mobilize electoral opposition, judges tend to be cautious in their sentences throughout their terms but become increasingly likely to "overpunish" as their reelection date approaches. In their words:

> If incumbent judges, like all officials, continuously reevaluate the balance between the value of office and implementing their own preference, the balance will shift increasingly toward the former as election approaches. Further, the fire-alarm nature of trial judge oversight suggests that because voters typically have access only to information about sentencing perceived as overly lenient, the judge's response will be *unidirectional convergence*. In other words, our theory predicts that judge will become more ***punitive***, not more congruent, over the course of their terms (250).

Huber and Gordon's theory implies a public whose opinions on punitiveness are both strong and pro-punishment. After all, if the public is divided on issues of punishment, there is little reason for a judge to to bolster her punitive credentials.

However, empirical evidence on attitudes toward criminal punishment stands in contrast to this assumption. Public attitudes toward punitiveness are not invariant. For example, Enns (2013) documents that American national attitudes toward punitiveness have varied widely over the past quarter-century. The trends are equally important at the state-level, where Enns (2013) demonstrates substantial cross-sectional variation in American's attitudes toward criminal punishment. While a majority of the public does favor strong punitive policies, individuals (and constituencies) vary in the extent to which they are supportive of these policies.

Judges in the American states are retained through a variety of elective and appointive methods. Among those judges who are elected, they may run in contestable partisan elections, contestable nonpartisan elections, or uncontestable nonpartisan retention elections. These different electoral types produce different incentives for judges. In uncontestable retention elections, judges are nearly assured of reelection unless some organized interest group mobilizes

against them. On the other hand, judges who run in contestable elections run the risk that a challenger will emerge and attempt to paint them as not tough enough on crime. Hence, as Gordon and Huber (2007) show, contestable elections may be associated with more punitive outcomes than retention elections. Moreover, Caldarone, Canes-Wrone and Clark (2009) argue that the lack of party labels on the ballot encourages judges who run in nonpartisan election to exaggerate their congruence with public opinion when compared to those judges who will be aided by party labels on the ballot. Hence, nonpartisan elections should lead to more congruence with public opinion than partisan elections.

To this end, the research demonstrates that public opinion has a conditional effect on judicial decisions, one reliant on the type of institutional mechanism used to determine whether a judge stays in office. Caldarone, Canes-Wrone and Clark (2009) demonstrate that state supreme court judges who run in partisan elections are less likely to reflect public opinion in their decisions than those who run in contestable nonpartisan elections. Similarly, Canes-Wrone, Clark and Kwang Park (2012) study the effects of issue-specific public opinion on judicial decisionmaking in death penalty cases and find that, though they can never face an opponent on election day, judges retained through yes-or-no retention elections tend to reflect public opinion in their decisionmaking, perhaps doing so even better than judges who run in partisan elections. Most recently, Canes-Wrone, Clark, and Kelly (2014), again studying death penalty cases, find that nonpartisan elections–both those that feature challengers and those that do not–encourage judges to follow public opinion to a higher degree than partisan elections do.

These findings stand in contrast with results at the trial court level. In a unique and important study, Gordon and Huber (2007) rely on the unique structure of the Kansas judiciary: roughly half of its trial court judges are retained using contestable partisan elections while the other half of its judges keep their jobs based on uncontestable retention elections. Gordon and Huber (2007) argue that, if the public is overwhelmingly punitive and contestable judicial elections drive judges to follow public opinion in their decisions, then those judges who are retained through contestable partisan elections should sentence more punitively than their colleagues

who will never face a challenger on election day. This is exactly what their results show: those Kansas judges who run in partisan elections sentence more punitively than the judges who face uncontestable elections.

The Canes-Wrone and Clark studies present a direct contradiction to the Gordon and Huber evidence from Kansas. The Canes-Wrone and Clark studies, including those that study the quintessential criminal law issue—the death penalty—suggest that nonpartisan elections induce more congruence than partisan elections. The Gordon and Huber evidence, built on the assumption of an invariantly punitive public, suggest the opposite.

How do we explain this contradiction? Most obviously, the appellate court studies model directly public opinion, examining the effects of the multiplicative interaction between judicial retention institution and state-level public opinion; the Gordon and Huber study relies on the assumption that public opinion on crime is highly punitive and invariant. Perhaps if one directly measured public opinion toward crime, using a model specification similar to that used by Canes-Wrone, Clark and Kelly (2014), the contridiction would be resolved.

At the same time, cross-sectional studies, like Canes-Wrone, Clark and Kelly (2014) must compare case outcomes across jurisdictional boundaries, therefore comparing outcomes across different state constitutions, bodies of precedent, and legal cultures. Methods of judicial selection were adopted in the United States due to a variety of extralegal concerns (Driscoll and Nelson 2013). If these concerns, which were often formalized in state constitutional conventions (Shugerman 2010, 2012) are correlated with the state constitutional provisions to which judges today respond or to a state's legal culture, then perhaps what appears to be the role of public opinion is simply due to these spurious factors.

Likewise, appellate court studies must grapple with the fact that judges on a collegial court influence each other (Maltzman, Spriggs II and Wahlbeck 2000), yet judicial opinions commonly have a single author. What we observe as a judge's vote, therefore, includes both her own rationale as well as some influence by her colleague. Because judges on most state appellate courts have different electoral clocks, their own sensitivity to public opinion may vary over time.

A variety of studies have demonstrated that judges change their behavior as their election or reappointment approaches (Berdejó and Yuchtman 2013; Huber and Gordon 2004), acting more punitively as the the date of their reselection approaches. Thus, the literature suggests both that public opinion affects judges in two ways: the method of selection affects their degree of congruence with public opinion while their electoral clock creates variation in their behavior throughout their term.

Hypotheses

Crime is an extraordinarily salient issue in judicial elections, and judges should therefore be particularly congruent with public opinion regarding punitiveness. Yet, the literature reports a contradiction between the effect of electoral institutions on judicial decisionmaking: appellate court studies indicate that nonpartisan elections lead to outcomes more congruent with public opinion, while Gordon and Huber's (2007) study implies the opposite about trial courts. However, the Gordon and Huber study does not model public opinion directly. To this end, I investigate the conditional effects of public opinion on judicial sentencing in Kansas, testing three hypotheses.

Constituency Opinions. Opinions and attitudes vary by individual and, by extension, by jurisdiction. Trial court judges in Kansas judges face a special hurdle: while they typically hear cases only in a single county, most judicial districts in Kansas span multiple counties. Thus, they have two different sets of constituents, in Fenno's (1978) sense: district constituents, who form the basis of their constituency but may or may not be part of the county whose cases the judge hears, and county constituents, which are those individuals who live in the county in which the judge hears cases. Given that counties within a single judicial district have varying opinions, meaning that district-level opinion diverges from county-level opinion, to which opinion should judges respond?

Of course, trial court elections are local elections. In trial court elections, candidates rarely run television advertisements, relying instead on newspaper and radio advertisements

and word of mouth. I expect that these latter, extremely local, considerations are particularly prominent in these elections because they are normally so low salience. To this end:

> *Hypothesis 1*: Judges should be more congruent with county-level than district-level opinion.

The Conditional Effect of Institutions. Canes-Wrone, Clark, and Kelly (2014) imply that judges retained through partisan elections should sentence less punitively, on average, than those judges who are retained via uncontestable elections. Yet, Gordon and Huber (2007) find the opposite to be true in Kansas. To this end, I follow Canes-Wrone, Clark, and Kelly in suggesting that

> *Hypothesis 2*: Judges who face nonpartisan retention elections should sentence more punitively than those who face partisan elections.

Electoral Proximity. In both of their studies, Gordon and Huber (2007; Huber and Gordon 2004) have demonstrated that trial court judges become more punitive as the date of their election approaches out of a desire to deter organized opposition, even finding that the effect of electoral proximity is conditional on the type of election facing a judge. Berdejó and Yuchtman (2013) demonstrate that judges in Washington give sentences that are 10% longer at the end of their terms than they are at the beginning. Canes-Wrone, Clark, and Kelly (2014) note a similar effect: state supreme court judges become more likely to vote to impose a death sentence as the date of their reselection approaches, though they do not examine whether this effect is conditional on either public opinion or judicial retention mechanism.

However, no study addresses the underlying assumption in these findings: judges sentence more punitively as their reselection date approaches *because* the public wants them to be more punitive. If this is the mechanism that underlies the proximity effect, then the magnititude of the proximity effect should be conditioned both by the type of reselection mechanism (since different mechanisms affect the likelihood of a challenger emerging and of eventual defeat) and

on the constituency's punitiveness (with those judges who have the most punitive constituencies becoming the most punitive as reselection approaches) Thus, I suspect that:

> *Hypothesis 3*: Electoral proximity has an effect on the length of judicial sentences conditional on the type of retention mechanism and constituency punitiveness. Judges who run in partisan elections with the most punitive constituencies should have the largest proximity effect.

Data

The data come from two places. First, I introduce a new extension to multilevel regression and post-stratification (Lax and Phillips 2009) which enables the estimation of county-level public opinion. The data for the public opinion estimates come from all available Gallup polls in the Roper iPoll archive from 1995-2010 and the national cross-sections of the 2000 Annenberg National Election Study. Second, I requested and received the Gordon and Huber (2007) data from the Kansas Sentencing Commission along with the other variables they used in the Gordon and Huber (2007) study to enhance comparability.

Dependent Variable

The outcome variable in the analysis is the number of months of prison assigned to the defendant at sentencing. Because the outcome is heavily skewed, I follow Gordon and Huber (2007) and use as the dependent variable $\ln(1 + \text{Assigned Prison Time in Months})$.

Measuring Local-Level Punitiveness

Scholars have long struggled to measure subnational public opinion (Erikson, Wright and McIver 1993; Brace et al. 2002). The state-of-the-art technique, multilevel regression and post-stratification (MRP) uses a two-step process to estimate subnational opinion from nationally representative surveys. In the multilevel regression stage, scholars estimate the probability that groups of people (e.g. a Latina with a high school education) in a particular jurisdiction support a particular policy. Here, the estimation stage includes covariates for race/ethnicity (white,

black, latino), education level (less than high school education, high school education, some college, college graduate), gender, and presidential vote share. Then, in the poststratification stage, census data is used to determine the proportion of the population within a jurisdiction of each type and to determine the average opinion. This procedure has been used and validated at the state (Lax and Phillips 2009) and congressional district (Warshaw and Rodden 2012) levels.

This procedure is difficult to translate to the county level because few surveys either record or make publicly available the county of residence for their respondents. Moreover, even if surveys contained the county of residence for their respondents, the unequal distribution of the population by county means that many most counties in the country would lack a respondent (let alone multiple respondents) from which accurate county-level effects can be estimated.

I sidestep this issue by performing the multilevel regression stage of the MRP procedure at the state level, obtaining estimates for each demographic type, each region, and each state. Then, I use the state- and region-level fixed effects for Kansas and poststratify at the county level (and the judicial district level). The resulting estimates are county- and district-level estimates of public opinion.[4]

To validate the measurement strategy, I compared the public opinion estimates of opinion on the death penalty from this proposed measurement strategy (employing a single nationally-representative Gallup survey) with the public opinion estimates from two representative single-state surveys (in which traditional MRP can be used, treating counties as states): the California Field Poll and the University of North Carolina's Odum Institute's 1999 Carolina Poll.[5] Figure shows the relationship between the estimates for both states; the top two plots show the estimates from the California polls while the bottom plots show the estimates from North Carolina. The left-hand plots include only those respondents who held an opinion on the death penalty; the right-hand plots include all respondents.

[4]Importantly, this measurement technique requires the assumption that, within a state, additional county- or district-level intercepts are not needed to soak up cross-county variation not covered by the demographic variables.

[5]There is a slight difference in wording in the California survey. Whereas the Gallup and North Carolina surveys use the same wording (asking whether or not respondents are in favor of the death penalty for individuals convicted of murder), the California poll asks respondents whether they believe the death penalty should be abolished.

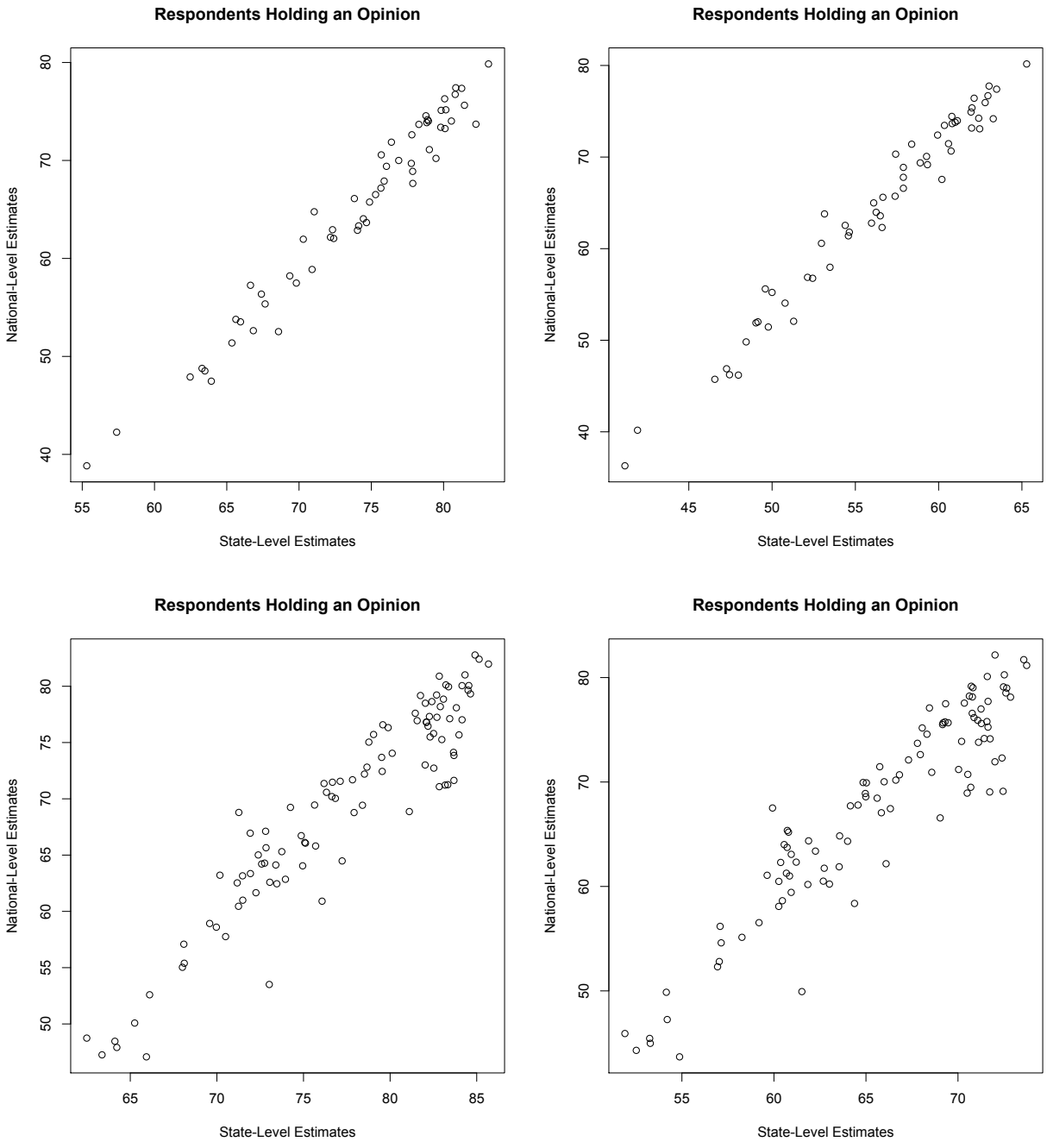

Figure 1: Public Opinion Estimates. These plots show the relationship between opinion on the death penalty using a single-state survey and a nationally-representative survey. The top two plots show the estimates from the California polls while the bottom plots show the estimates from North Carolina.

The correlations between the estimates derived from the nationally-representative survey and from the single-state surveys provide strong evidence of the validity of the proposed estimation technique. For North Carolina, the measures including all respondents correlate at $r = 0.93$ while the measures including only those respondents with an opinion correlate at $r = 0.94$. For California, the correlations are nearly perfect: $r = 0.98$ when including only those respondents who held an opinion and $r = 0.99$ when including all respondents.[6]

Having provided some evidence of the validity of the estimation technique, we now turn to the problem of estimating opinion on punitiveness. Recently, Enns has proposed measures of punitiveness at the national (2014) and state (2013) levels. Enns's measurement strategy relies on obtaining public opinion on a number of issues related to criminal justice and crime and then combining those measures into a single series. I take a similar approach, combining ten questions: all of the questions from Gallup surveys used by Enns asked between 1995 and 2010 as well as all questions in the 2000 Annenberg National Election Studies related to crime and criminal justice. Exact question wordings are found in Appendix A.

County-Level Correlations

	Death Penalty	Strict Gun Policy	Own Gun	Miranda	Police Conf.	Crim. Just. Conf.	Underpun.	Drug	A'burg Death Penalty	A'burg License Gun
Death Penalty	1.00									
Strict Gun	-0.01	1.00								
Own Gun	0.89	-0.44	1.00							
Miranda	0.91	0.39	0.66	1.00						
Pol. Conf.	0.81	0.57	0.49	0.98	1.00					
CJ Conf.	0.69	0.71	0.32	0.92	0.98	1.00				
Underpunish	0.82	0.55	0.51	0.98	0.99	0.97	1.00			
Drug	0.70	0.70	0.33	0.93	0.98	1.00	0.98	1.00		
A'burg Death P.	0.95	0.29	0.73	0.99	0.95	0.88	0.96	0.88	1.00	
A'burg License Gun	0.24	0.96	-0.19	0.61	0.76	0.87	0.74	0.86	0.52	1.00

Table 1: Correlation matrices demonstrating the relationship between all indicators of public opinion at the county level. There are 105 counties in Kansas.

Table 1 shows the correlations among the county-level opinion estimates from each of the ten questions. The high correlations provide convincing evidence that the ten indicators relate

[6]The graphs do indicate some level of bias between the public opinion estimates estimated using the two different procedures; the new procedure seems to underestimate public opinion on crime. While further analyses need to determine the extent to which any bias that the method produces is consequential, the measure is appropriate for these analyses because they are being used in regression analyses and therefore the relative distance among observations, rather than the mean, is important for the overall regression estimates.

to the same concept.

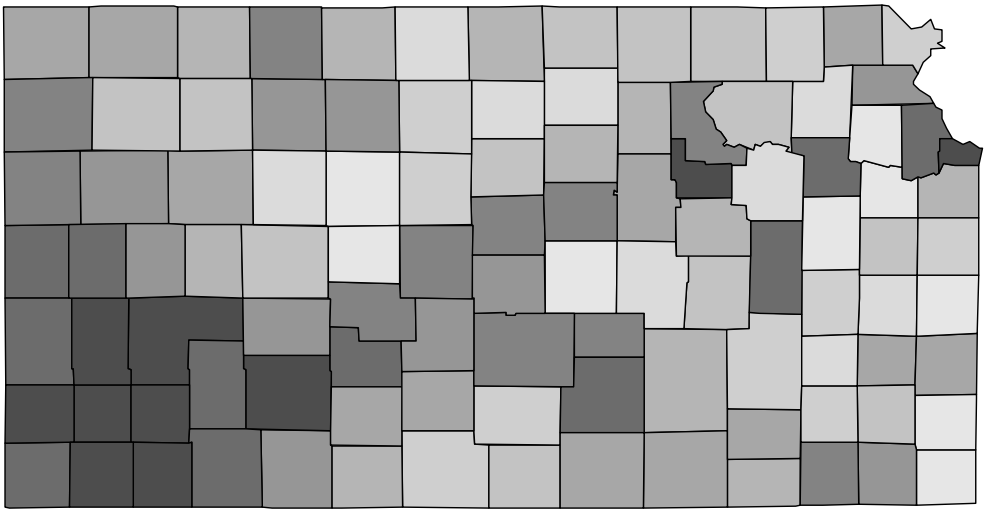

Figure 2: Public Opinion Estimates. The colors on the map are on a gradient; darker colors indicate more punitive attitudes.

The next step is to combine the ten indicators into a single indicator of punitiveness. Because Enns's measure of punitiveness is dynamic, Enns employs Stimson's Wcalc algorithm for this purpose. Because my measure of punitiveness is not dynamic, I use confirmatory factor analysis. Further analyses indicate that punitiveness is unidimensional and a single factor is appropriate. While the two Gallup indicators relating to gun policy load only moderately onto the single factor, the rest of the indicators load very highly on the factor. Because the factors scores drawn from measures of punitiveness with and without those two Gallup gun policy indicators correlate above $r = 0.98$, I use the measure of punitiveness which includes all of the indicators. Figure 2 shows the distribution of estimated county-level punitiveness across the state. On the figure, darker colors indicate more punitive opinions.

How valid is the measure of punitiveness? I investigated the validity of the measure by examining the correlations between the measure and presidential vote share. Interestingly, puni-

tiveness is unrelated to presidential vote share, with $r = 0.01$. However, this relationship is not unexpected; the correlation between a 2000 Annenberg National Election Studies item asking about whether respondents believe underpunishment is a problem is similarly uncorrelated with the share of the vote won by the Democratic presidential candidate: $r = -0.02$. Other indicators of punitiveness correlate as expected with presidential vote share. The relationship between opinion on the death penalty and vote share is $r = -0.54$ and between gun ownership and vote share is $r = 0.85$.

Having measured public opinion at the constituency level, one important preliminary question concerns the relationship between constituency-level punitiveness and judicial selection method. Judicial selection methods in Kansas, like all institutions, are not randomly assigned, so one may wonder whether more punitive districts tend to have a particular selection method. Of the 31 judicial districts in Kansas, 14 use partisan elections and 17 use retention elections to determine whether a judge will stay in office. 48 of the state's 105 counties use partisan judicial elections. There is no relationship between method of judicial retention and the punitiveness at the district ($p = 0.12$) or the county ($p = 0.97$) levels. Yet, there is a relationship between Gore's vote share in a district and whether that district uses partisan elections, with Gore doing worse in districts that use partisan judicial elections ($p < 0.00$). This suggests that, while punitiveness is not related to election type, election type is related to district ideology. More conservative districts are more likely to utilize contestable partisan judicial elections. This finding is important. If one of the election types—contestable partisan elections—tend to be used in districts that are also more punitive, then what looks to be an institutional effect may simply be due to judges reflecting public opinion, albeit one that varies across districts.

Of course, judicial districts vary in the number of cases they hear in a given year. Thus, when the data are expanded to the case level, there is a statistically significant relationship between punitiveness and whether the county or district uses partisan judicial elections. In both cases, districts that use partisan elections are more punitive than those that use retention elections.

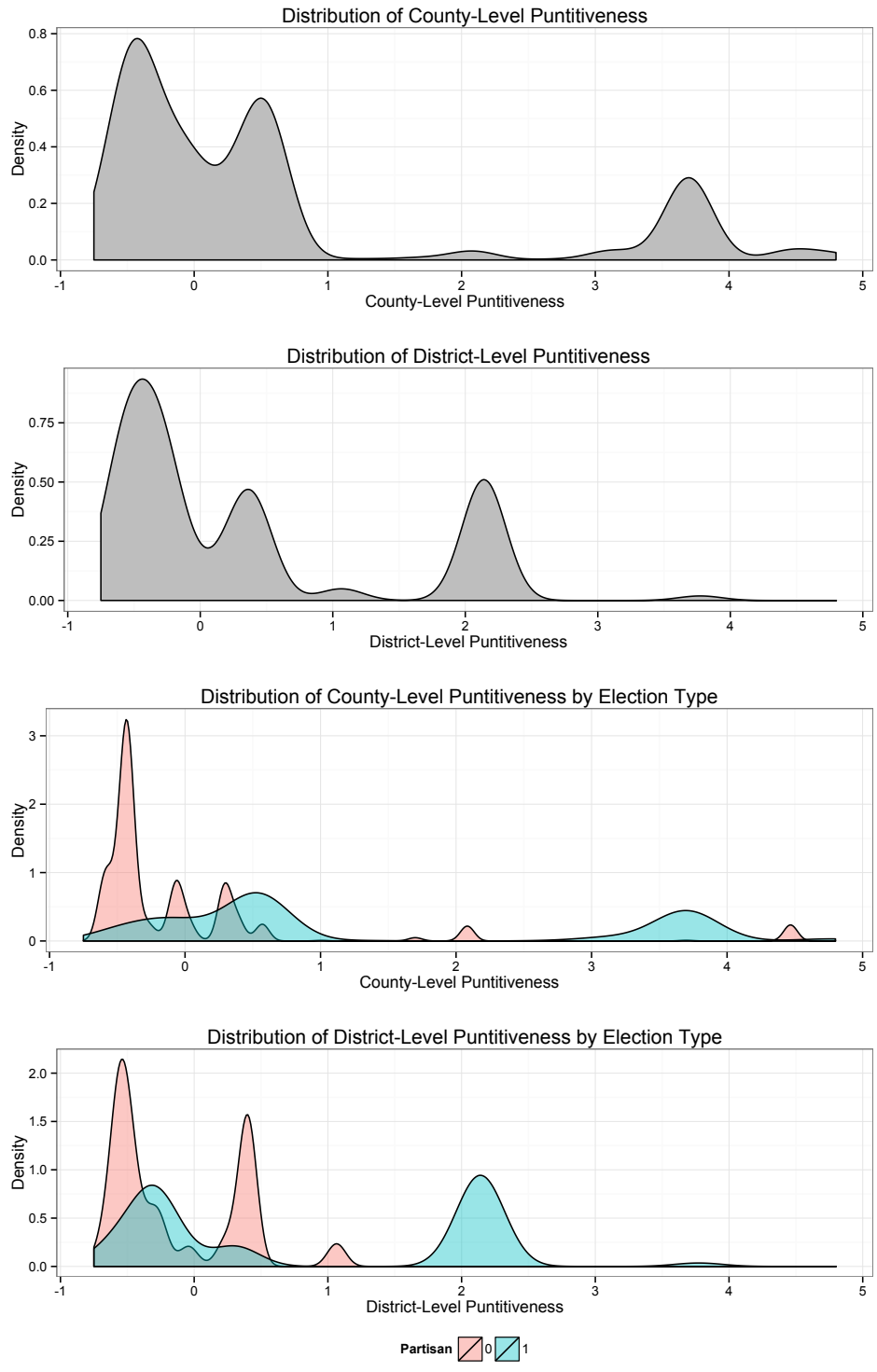

Figure 3: Density of punitiveness at the county- and district-levels. Though the factor analysis forces the measures to have mean zero and a standard deviation of one across the 31 judicial districts or the 105 counties, each county has a differential number of observations in the data. Thus, the resulting densities do not resemble normal distributions.

This is seen most clearly in Figure 3. This figure which plots the density of observed puntitiveness at the case level. The x-axes of all four panels are the same and demonstrate that the average punitiveness at the district level (0.35) is more moderate than the average at the county level (0.65). The bottom two panels of the figure separate observed punitiveness by election type and illustrate the clear difference in public opinion with regard to crime by election type. Whereas the average punitiveness for counties that use retention elections is -0.02, the average for those that use partisan elections is 1.33; at the district level, the average is -0.14 for retention election districts and 0.84 for partisan election districts.

Other Independent Variables

To ensure the comparability of the findings, all of the remaining variables in the study are measured the same as in Gordon and Huber (2007). The measure of electoral proximity is the same as Gordon and Huber's. The measure ranges from 0 (when the judge's new term begins or when he wins reelection on the general election day) to 1 (when the election is just days away) and resets to zero on the day of the filing deadline for the impending election when a judge in a contestable election learns that she will be unchallenged. The measure has a mean of 0.50 with a standard deviation of 0.26.

The model includes a number of control variables. These variables include the number of counts in the conviction, the type of crime (assault, criminal threat, robbery, sex crime, theft, burglary, crime carried out with a firearm), whether the sentence was reached by a jury or a plea bargain, whether counsel was private or appointed, and the presumptive prison sentence for the top count in the case (which is determined the the defendant's prior criminal history). I also include characteristics of the victim (government/law enforcement official, child) and characteristics of the victim (race, ethnicity, gender, age, sex offender status). In some models, I also include a set of district-specific control variables, including presidential turnout, presidential vote share, mean retention vote share, the proportion of the district that is nonwhite, the proportion urban, and the crime rate.

The models all include year fixed effects both to account for any time trends in the data

but also to account for prosecutorial discretion. As Nelson (2014) notes, prosecutors, like judges, may respond to public opinion. In the Kansas case, this concern is mitigated by the fact that all prosecutors have the same four-year electoral calendar while judges have staggered terms. Thus, the year fixed effects also soak up variation related to the prosecutor's electoral calendar.

Method

Gordon and Huber (2007) used Tobit to model the sentences handed down to defendants in Kansas. Here, for ease of interpretation, I present linear regressions, noting that the results from Tobit regressions are nearly identical. Because the same judge sentences multiple defendants in the dataset, the observations are not strictly independent; thus, I use heteroscedasticity-consistent standard errors, clustered on the judge, to account for this dependence.

Results

	No Interactions		With Interactions		Interactions + Controls	
	County	District	County	District	County	District
	(1)	(2)	(3)	(4)	(5)	(6)
Partisan Election	0.04	0.12*	0.05	0.12*	0.12*	0.11*
	(0.04)	(0.03)	(0.05)	(0.03)	(0.05)	(0.04)
County-Level Punitiveness	0.03*		0.06*		0.06*	
	(0.01)		(0.01)		(0.03)	
District-Level Punitiveness		-0.04		0.03		0.12*
		(0.02)		(0.03)		(0.05)
Partisan×County-Level			-0.04*		-0.05*	
			(0.02)		(0.02)	
Partisan×District-Level				-0.08		-0.12*
				(0.04)		(0.04)
ln(1+Presumptive Prison)	0.61*	0.61*	0.61*	0.61*	0.61*	0.61*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
First guideline regime	0.14*	0.14*	0.14*	0.14*	0.13*	0.14*
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Second guideline regime	0.09*	0.08*	0.09*	0.08*	0.08*	0.09*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Number of additional	0.08*	0.09*	0.08*	0.09*	0.09*	0.09*
counts in conviction	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Persistent sex offender	1.16*	1.16*	1.16*	1.16*	1.16*	1.17*
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.21)
Person felony committed	0.68*	0.68*	0.68*	0.68*	0.68*	0.69*
with a firearm	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Assault	-0.01	-0.00	-0.02	-0.00	-0.00	0.00
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Criminal Threat	-0.20*	-0.19*	-0.20*	-0.20*	-0.20*	-0.19*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)

Robbery	0.40*	0.41*	0.40*	0.41*	0.42*	0.42*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Sex crime	0.69*	0.70*	0.69*	0.70*	0.70*	0.70*
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Theft	-0.03	-0.02	-0.03	-0.03	-0.02	-0.02
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Burglary	0.02	0.03	0.02	0.03	0.03	0.03
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Victim was government	0.31*	0.30*	0.31*	0.30*	0.30*	0.30*
or law enforcement official	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Victim was a child	-0.25*	-0.26*	-0.25*	-0.26*	-0.25*	-0.26*
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Appointed counsel	0.27*	0.27*	0.27*	0.28*	0.26*	0.27*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Plea bargain (baseline is jury)	-0.72*	-0.72*	-0.72*	-0.72*	-0.72*	-0.72*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Defendant Nonwhite	0.01	0.03	0.02	0.03	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Defendant Male	0.27*	0.27*	0.27*	0.27*	0.26*	0.26*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Defendant Hispanic	0.05	0.08*	0.04	0.08*	0.05	0.06
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)
Defendant age	0.05*	0.05*	0.05*	0.05*	0.05*	0.05*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Defendant age, squared	-0.00*	-0.00*	-0.00*	-0.00*	-0.00*	-0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Presidential Turnout					0.50*	0.43*
					(0.24)	(0.20)
Gore Vote Proportion					0.12	-0.06
					(0.35)	(0.31)
Proportion nonwhite					0.32	0.92*
					(0.61)	(0.31)
Urban					-0.25*	-0.38*
					(0.10)	(0.12)
Crime Rate					0.00	0.00*
					(0.00)	(0.00)
Yes Retention Pct.					-0.10	0.14
					(0.73)	(0.69)
Constant					-0.64	-0.72
					(0.57)	(0.54)
Adjusted R^2	0.71	0.71	0.71	0.71	0.59	0.59
N	18141	18141	18141	18141	18141	18141

Table 2: The effects of public opinion about punitiveness on the sentences of Kansas judges. The models are linear regression, with $\ln(1 + \text{Assigned Prision Time in Months})$ as the dependent variable. The models all include fixed effects for year.

Table 2 provides the results of the regression analyses. The table contains three pairs of models. Models 1 and 2 regress punitiveness at the county- and district-level on the sentence

received by the defendants. Models 3 and 4 add the multiplicative interaction between punitiveness and election type to the model. Models 5 and 6 retain the multiplicative interaction while also adding a host of district-level control variables to the specification.

We begin by examining the effect of punitiveness at the simplest level: unconditional on election type, to what extent does public opinion on crime correlate with the sentences judges receive? Models 1 and 2 provide these results, which suggest different roles for county- and district-level public opinion. At the county level, election type has no independent effect on sentence length, controlling for county-level punitiveness; moreover, Model 1 suggests that, as the public's desire for harsher sentences increases, defendants receive longer sentences. This is not true at the district level, where public opinion is uncorrelated with sentence length.[7] However, in this model, the dichotomous indicator for partisan elections provides support for Gordon and Huber's (2007) suggestion that partisan elections are associated with longer sentences. Thus, there is some support for Hypothesis 1: judges seem more congruent with county, rather than district, level opinion, at least when opinion is examined only for its direct effect.

Of course, the more subtle—and more important—question is whether election type conditions the effect of public opinion on judges' sentences. Across Models 3-6, the answer to that question is yes in all but one model. The negative coefficient on the interaction terms suggest that the effect of partisan elections on sentence length decreases as the public becomes more punitive.

This is seen most clearly in Figure 4 which plots the marginal effect of partisan elections on sentence length across the range of punitiveness. Recall from Figure 3 that the range of district-level punitiveness is smaller than that of county-level punitiveness which is why the line on the bottom panel does not extend across the entire figure (the x-axes in the two figures are the same). The figures clearly show what the sign on the coefficients suggested: the punitive effect of partisan elections decreases as the public become more punitive and become indistinguishable from zero.

[7] This is further true when both county- and district-level opinion are included in the model together; county-level opinion is statistically significant while district-level opinion is not.

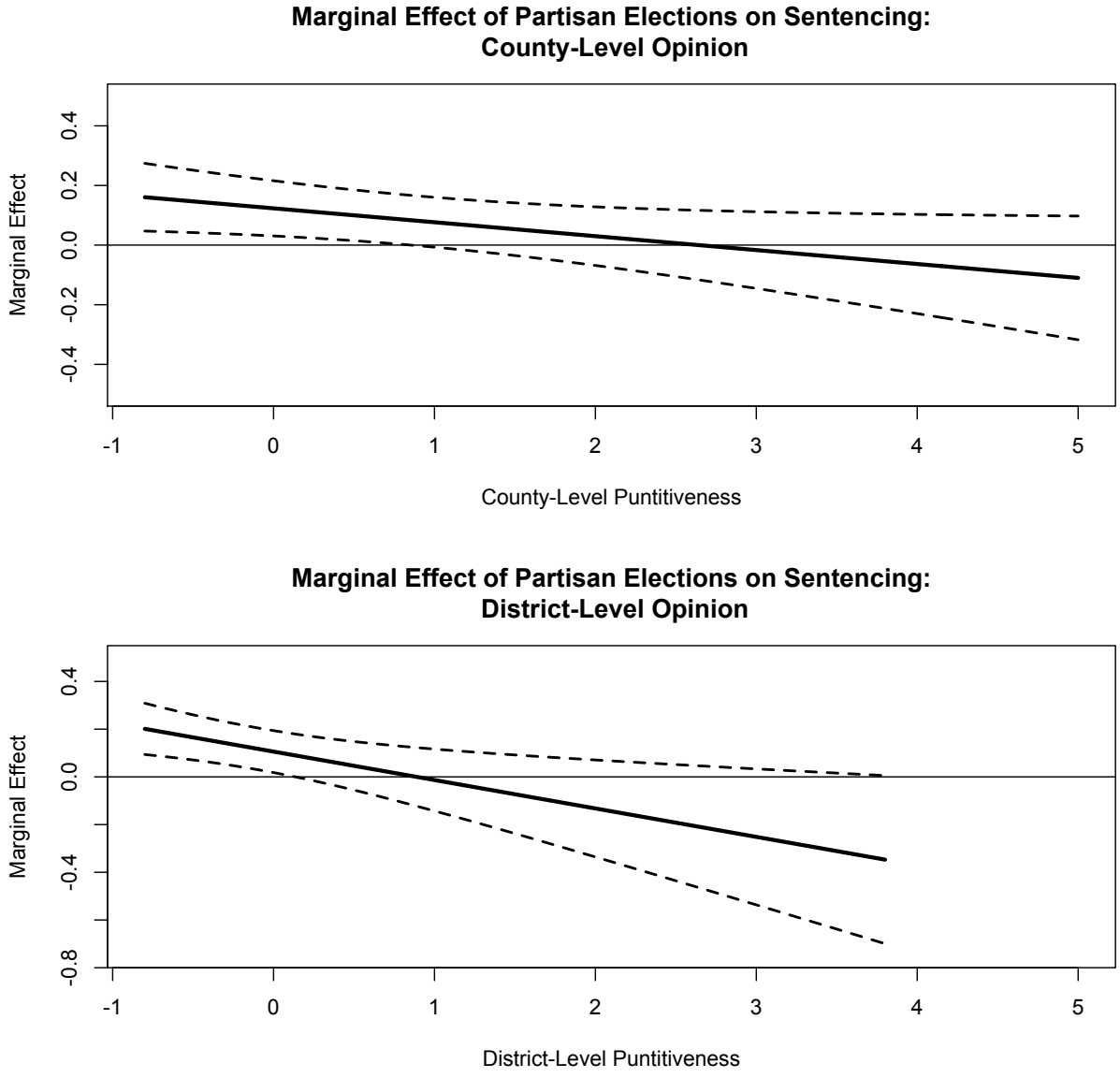

Figure 4: Marginal effect plots. These plots show the marginal effect of partisan (rather than retention) elections on the sentences assigned by Kansas judges across observed levels of punitiveness.

While the marginal effect appears from the graph to be statistically insignificant for most of the range of the variable, about 10% of counties and 25% of judicial districts fall into the range of values for which the effect of partisan elections is not statistically significant. Thus, these results indicate that, while partisan elections are associated with more punitive judgments, this is only the case when constituency opinion is not very punitive; in the most punitive jurisdictions, there is no difference in the sentencing tendencies of judges regardless of their retention method.

Figure 4 tells us how the difference in average sentences among partisan-elected and retention-elected judges varies over the range of public opinion. We can also examine the interaction term from the other direction to examine how the effect of public opinion on judicial sentences changes for each of the two election types. For both county- and district-level public opinion, judges who face uncontestable retention elections are *more* likely to follow public opinion than those who run in partisan elections. For retention-elected judges, the magnitude of the effect of punitiveness on sentences is 0.06 for county-level opinion and 0.12 for district-level public opinion. For those judges facing partisan elections, public opinion has no effect on their decisions, with coefficients of 0.01 for county-level public opinion and 0.00 for district-level public opinion. Thus, contrary to the findings of Canes-Wrone, Clark and Kelly (2014), it appears that Kansas trial court judges who face partisan elections do not reflect public opinion whereas judges who will never face a challenger are broadly reflective of public opinion in their sentences.

Table 3 expands the analysis by introducing electoral proximity. Models 1 and 2 begin with the model specifications from the final two models in Table 2, adding a direct effect for electoral proximity. The middle two models in the table add a two-way multiplicative interaction for proximity and election type while the final two models include a three-way interaction between election type, punitiveness, and electoral proximity.

The results of the model suggest that, contrary to the findings of Canes-Wrone, Clark, and Kelly (2014), electoral proximity has no discernible effect on judicial behavior after controlling for the conditional effect of election type on public opinion. There is no evidence that judges

	Direct Effect		Two-Way Interactions		Three-Way Interaction	
	County	District	County	District	County	District
	(1)	(2)	(3)	(4)	(5)	(6)
Partisan Election	0.12*	0.11*	0.04	0.02	0.05	0.01
	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Electoral Proximity	0.01	0.01	-0.07	-0.07	-0.07	-0.07
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
County-Level Punitiveness	0.06*		0.06*		0.09	
	(0.03)		(0.03)		(0.05)	
Partisan×County-Level	-0.05*		-0.05*		-0.08	
	(0.02)		(0.02)		(0.05)	
District-Level Punitiveness		0.12*		0.12*		0.11
		(0.05)		(0.05)		(0.07)
Partisan×District-Level		-0.12*		-0.12*		-0.09
		(0.04)		(0.04)		(0.07)
Partisan×Proximity			0.16*	0.16*	0.14	0.19*
			(0.07)	(0.07)	(0.08)	(0.07)
County-Level×Proximity					-0.04	
					(0.07)	
Partisan×County-Level×Proximity					0.06	
					(0.08)	
District-Level×Proximity						0.02
						(0.09)
Partisan×District-Level×Proximity						-0.05
						(0.10)
Adjusted R^2	0.59	0.59	0.59	0.59	0.59	0.59
N	18141	18141	18141	18141	18141	18141

Table 3: The Role of Electoral Proximity. The models are linear regression, with $\ln(1 +$ Assigned Prision Time in Months$)$ as the dependent variable. The models include year fixed effects and all other variables shown in models 5 and 6 in the previous table.

tend to sentence more punitively as the date of their election approaches.

Expanding the model to account for the fact that electoral proximity may have differential effect based on the type of election faced by a judge, there is some evidence that judges who face partisan elections do sentence more punitively as the date of their election approaches. Importantly this effect persists after controlling for county- or district-level public opinion and its interaction with election type. However, the effects dissipate once public opinion, electoral proximity, and election type are all interacted with each other.

Discussion and Conclusion

The major conclusion from the analysis is that, unlike at the state supreme court level, constituency-level pressures appear to exert a greater pull on judges—at least those in Kansas—who face contestable partisan elections rather than uncontestable retention elections. Whereas Canes-Wrone, Clark and Kelly (2014) found that judges in partisan election systems were less

likely than judges who face uncontestable retention elections to vote in favor of the death penalty, this analysis suggests the opposite. For most levels of constituency-level punitiveness, particularly at the county level, judges who face contestable elections do, as Gordon and Huber (2007) argued, sentence more punitively than their counterparts who will not face a challenger at the polls. Importantly—and what separates this study from Gordon and Huber's analysis—is the fact that this finding holds even when the possible conditional effects of constituency-level public opinion on judicial decisionmaking are considered.

Moreover, this analysis suggests, again contrary to Canes-Wrone, Clark and Kelly (2014) that retention-elected judges are *more* reflective of public opinion that judges who run in partisan elections. Whereas retention-elected judges with more punitive constituencies do, on average, give harsher sentences than their counterparts with less punitive constituencies, the same is not true for judges who face contestable partisan elections. For these judges, there is no evidence that the punitiveness of their sentence changes based on their constituents' level of punitiveness. On the one hand, these results fit well with the findings of Nelson (2014), who found that Colorado judges who face retention elections are highly congruent with public opinion, even changing their behavior in the face of new evidence about constituency preferences.

As this project moves forward, more work needs to be done on the measure of punitiveness. As noted earlier, Enns (2013) uses the full list of questions relating to punitiveness found in the General Social Survey, the ANES, and Roper's iPoll archive, while the analysis here uses only Gallup and Annenberg questions. More questions and more validation work—particularly regarding the district-level measure of public opinion where the correlations among indicators are extraordinarily high—is necessary.

Yet, the results raise two big questions for further research. First, why do local-level partisan-elected judges fail to respond to varying levels of public opinion while their counterparts who face retention elections do? Local level public opinion is difficult to measure, and one possibility may be that, for some reason, those judges who face partisan elections simply have a more difficult time ascertaining their constituents' opinions than those who face retention elec-

tions. By this theory, incongruence with public opinion is due to lack of information rather than willful disregard. A more plausible theory is based on the dynamics of these types of campaigns and reconciles the result that partisan elected judges tend to sentence more harshly on average. Perhaps judges who face partisan elections, as Gordon and Huber (2007) suggest, are wary of drawing a challenger so they always sentence on the harsh side of the distribution in order to mitigate their eventual opponent's ability to draw a challenger. Thus, these judges are not reflecting variation in public opinion because they are always trying to sentence more harshly. On the other hand, judges who face uncontestable retention elections worry about drawing an organized opposition campaign—something more difficult to put together than a challenger in a partisan election—and so they are able to be more in tune to variation in their constituents' preferences than their colleagues in jurisdictions that use partisan elections.

Second, why is there a contradiction regarding the effects of judicial retention institutions at different levels of the judicial hierarchy? One likely answer to this question involves the difference in the salience of campaigns at the appellate and trial court levels. Whereas most recent contestable state supreme court elections involve challengers (Hall 2001), Nelson, Caufield and Martin (2013) shows that trial court judges running in contestable elections face a challenger—at either the primary or general election stages of the electoral process—in less than 25% of judicial contests. Similarly, appellate court elections are much more likely to feature television advertising, huge amounts of campaign spending, or even widespread news coverage. Given that even retention elections at the appellate court level can often involve high profile anti-retention campaigns (Canes-Wrone, Clark and Kwang Park 2012), the electoral incentives for appellate and trial court judges are very different. Future work must investigate how these incentives differ across levels of court to broaden our theories about how and why retention institutions affect congruence with public opinion.

References

Bailey, Michael A. and Forrest Maltzman. 2011. *The Constrained Court: Law, Politics, and the Decisions Justices Make.* Princeton University Press.

Baum, Lawrence. 2003. "Judicial Elections and Judicial Independence: The Voter's Perspective." *Ohio State Law Journal* 64:13–42.

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

Blume, John and Theodore Eisenberg. 1998. "Judicial Politics, Death Penalty Appeals, and Case Selection: An Empirical Study." *Southern California Law Review* 72:465–503.

Brace, Paul, Kellie Sims-Butler, Kevin Arceneaux and Martin Johnson. 2002. "Public Opinion in the American States: New Perspectives Using National Survey Data." *American Journal of Political Science* 46:173–189.

Bright, Stephen B. and Patrick J. Keenan. 1995. "Judges and the Politics of Death: Deciding Between the Bill of Rights and the Next Election in Capital Cases." *Boston University Law Review* 75:759–835.

Caldarone, Richard P., Brandice Canes-Wrone and Tom S. Clark. 2009. "Partisan Labels and Democratic Accountability: An Analysis of State Supreme Court Abortion Decisions." *Journal of Politics* 71:560–573.

Canes-Wrone, Brandice, Tom S. Clark and Jason P. Kelly. 2014. "Judicial Selection and Death Penalty Decisions." *American Political Science Review* 108(1):23–39.

Canes-Wrone, Brandice, Tom S. Clark and Jee Kwang Park. 2012. "Judicial Independence and Retention Elections." *Journal of Law, Economics, & Organization* 28:211–234.

Champagne, Anthony. 2001. "Television Ads in Judicial Campaigns." *Indiana Law Review* 35:669–689.

Driscoll, Amanda and Michael J. Nelson. 2013. "The Political Origins of Judicial Elections: Evidence from the United States and Bolivia." *Judicature* 96:151–160.

Enns, Peter K. 2013. "Punitive Politics in the U.S. States." Paper Presented at the Annual Meeting of the Midwest Political Science Association.

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

Erikson, Robert S., Gerald C. Wright and John P. McIver. 1993. *Statehouse Democracy.* Cambridge University Press.

Fenno, Richard F. 1978. *Home Style: House Members in their Districts.* Little Brown.

Gordon, Stanford C. and Gregory A. Huber. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2:107=138.

Hall, Melinda Gann. 2001. "State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform." *American Political Science Review* 95:315–330.

Huber, Gregory A. and Stanford C. Gordon. 2004. "Accountability and Coercion: Is Justice Blind when it Runs for Office?" *American Journal of Political Science* 48:247–263.

Lax, Jeffrey R. and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1):107–121.

Linde, Hans A. 1987. "Elective Judges: Some Comparative Comments." *Southern California Law Review* 61:1995–2006.

Maltzman, Forrest, James F Spriggs II and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game.* Cambridge University Press.

Nelson, Michael J. 2014. "Representative Justice? Public Opinion and the Criminal Justice System." *Journal of Law and Courts* 2:117–152.

Nelson, Michael J., Rachel Paine Caufield and Andrew D. Martin. 2013. "OH, MI: On Empirical Examinations of Judicial Elections." *State Politics & Policy Quarterly* 13:495–511.

Shugerman, Jed. 2012. *The People's Courts.* Harvard University Press.

Shugerman, Jed Handelsman. 2010. "Economic Crisis and the Rise of Judicial Elections and Judicial Review." *Harvard Law Review* 123:1061–1150.

Warshaw, Christopher and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74:203–19.

Appendix

This appendix provides the specific question wording for the ten indicators used to construct the measure of punitiveness:

- Death Penalty (Gallup): "Are you in favor of the death penalty for a person convicted of murder?"
- Strict Gun (Gallup): "Would you like to see gun laws in this country made more strict, less strict, or remain as they are?" and "In general, do you feel that the laws covering the sale of firearms should be made more strict, less strict, or kept as they are now?"
- Own Gun (Gallup): "Do you personally own a handgun, rifle, shotgun or any other kind of firearm?" and "Do you have a gun in your home?"
- Miranda (Gallup): "Do you think confessions obtained from defendants who were not read their constitutional rights when they were arrested should or should not be admissible in trial?"

- Police Confidence (Gallup): "Now I am going to read you a list of institutions in American society. Please tell me how much confidence you, yourself, have in each one – a great deal, quite a lot, some, or very little? The Police"
- Criminal Justice Confidence (Gallup): "Now I am going to read you a list of institutions in American society. Please tell me how much confidence you, yourself, have in each one – a great deal, quite a lot, some, or very little? The Criminal Justice System"
- Underpunish (Annenberg): "The number of criminals who are not punished enough—is this an extremely serious problem, serious, not too serious or not a problem at all?"
- Drug (Annenberg): "The amount of illegal drug use—is this an extremely serious problem, serious, not too serious or not a problem at all?"
- Death Penalty (Annenberg): 'Do you personally favor or oppose the death penalty for some crimes?"
- Own Gun (Annenberg): "Do you personally favor or oppose requiring a license for a person to buy a handgun?"