THE EFFECT OF ELECTORAL COMPETITIVENESS ON INCUMBENT BEHAVIOR

Sanford Clark Gordon*　　Gregory Huber
New York University　　Yale University

Abstract

What is the marginal effect of competitiveness on the power of electoral incentives? Addressing this question empirically is difficult because challenges to incumbents are endogenous to their behavior in office. To overcome this obstacle, we exploit a unique feature of Kansas courts: fourteen districts employ partisan elections to select judges, while seventeen employ noncompetitive retention elections. In the latter, therefore, challengers are ruled out. We find judges in partisan systems sentence more severely than those in retention systems. Additional tests attribute this to the incentive effects of potential competition, rather than the selection of more punitive judges in partisan districts.

---
*Corresponding Author

1 Introduction

In empirical studies of the agency relationship between voters and elected officials, it has been well established that the behavior of the latter seems in many cases to conform to the demands of the former (e.g. Besley and Coate 2003; Bartels 1991; Miller and Stokes 1963). This conformity has been explained with reference to the success of voters at selecting likeminded officials and by the incentive effects of periodic review. To the extent that incumbents alter their behavior in anticipation of possible sanction at the polls, a puzzle emerges: To what extent does the competitiveness of the electoral environment increase the power of these incentives? This question has taken on particular importance in the American political context, where public officials in numerous positions frequently run for reelection unopposed or face only token opposition (Cox and Katz 1996).

The absence of serious challengers may be said to contribute to an erosion of democratic accountability. At the same time, however, this absence (and the resulting lopsided margins by which incumbents often retain office) may reflect incumbent compliance with public demands rather than autonomy from them. In addition to serving as alternatives for voters, serious challengers in competitive elections can audit incumbents and provide information to voters about the qualifications of incumbents or their behavior in office. The mere threat that a challenger could play these roles might be sufficient to alter incumbent behavior. Such a response would deny challengers a strong case to make to voters for the incumbent's replacement, or deter them from running altogether. In this regard, the *potential* for competition could improve incumbent compliance because of the roles that challengers *might* play, but, in the vast majority of circumstances, never get the chance to.

A central empirical implication of this argument is that incumbents should be more responsive to public demands when the potential for electoral competition is higher. Unfortunately, isolating the effect of competitiveness empirically is difficult because challenger viability – whether in primary or general election contests – is itself endogenous to incumbent performance. In other words, the spirited electoral contests that officeholders fear are the very ones in which a challenger has a case to make to voters about an incumbent's failures.

To overcome this problem, we rely on a unique feature of district court judicial elections in Kansas. There, fourteen judicial districts employ partisan competitive elections to select judges, while seventeen employ gubernatorial appointment and non-competitive retention elections. The retention districts provide an appropriate baseline of comparison: Although judges in those districts must face the voters, by law, *there can be no challengers.* Comparing officials in a single state permits us to hold constant numerous potential confounding factors, most importantly the legal environment in which those officials operate.

We derive predictions concerning the relationship between the sentencing behavior of incumbent judges and the electoral system in which they operate. Employing data on felony convictions in Kansas and several econometric approaches, we demonstrate that judges in partisan competitive systems sentence significantly more punitively than those in retention systems. Our identification strategy subsequently exploits variation in judges' electoral calendars to demonstrate that the effect of the electoral system is more pronounced in altering judges' incentives than in changing who is selected to office. Our findings demonstrate that the potential for competition from challengers substantially enhances incumbent attention to voter demands, while providing insight into both the functioning of electoral systems generally and the operation of the criminal justice system in particular.

2 Elections, Accountability, and the Role of Challengers

2.1 Selection and Incentives

Broadly speaking, there are two mechanisms by which elections might produce faithful representation on the part of elected officials. The first is selection. Ideally, competitive elections allow voters to choose candidates whose preferences most closely mirror their own (Downs 1957; Fearon 1999). In the selection account, the presence of challengers facilitates a closer match between voters and their representatives through the provision of alternatives. The second mechanism is the incentive effect of elections (Barro 1973; Ferejohn 1986). Even those incumbents who do not share their constituents' preferences or possess strong qualifications may nonetheless behave faithfully or work hard if their failure to do so will result in their subsequent defeat at the polls.[1]

2.2 The Functions of Challengers

The threat a challenger poses to an incumbent plays a critical role in structuring incumbent incentives. If voter preferences are heterogeneous, the presence of an additional choice on the ballot may be sufficient to reduce the vote share of an incumbent with an otherwise strong reputation. Challengers also play an informational role in elections. They are strongly motivated to audit incumbent performance and, when possible, publicize actual or perceived missteps (Arnold 1993). In this respect, challengers play a "fire alarm" role (McCubbins and Schwartz 1984), monitoring incumbents on behalf of voters and reporting evidence of shirking. This role complements the informational function of the media, constituency groups, and grassroots organizations, as attention to incumbent behavior is far more intense during competitive election seasons (Arnold 2004; Alvarez 1998). If entering a race entails opportunity costs for a challenger, the very fact that he or she has stepped into the ring may signal to voters that the incumbent is deficient in some regard (Gordon, Huber, and Landa 2006).

What should we expect to observe if incumbents alter their behavior in anticipation of this threat? First, we should frequently see races lacking a viable challenger. Viability is often a consequence of the incumbent's vulnerability, which he or she will take pains to minimize.[2] In the absence of viable challengers, incumbents should secure reelection by wide margins. Second, we should expect voters to be ignorant much of the time, because their knowledge of incumbent behavior is endogenous to campaigning by challengers (and the subsequent attention paid by the media).

In light of these difficulties, efforts to ascertain the incentive effect of electoral competition using data on voter knowledge and election outcomes will likely suffer from simultaneity bias. One might posit, for example, that a relationship between competition and compliance implies

---

[1]Of course, faithful agency by politicians to their constituents, while individually rational, might lead to collectively suboptimal results (e.g. Mayhew 1974; Maskin and Tirole 2004). We return to this general point in the conclusion.

[2]When we do observe viable challenges, it is likely due to unanticipated shocks that make incumbents vulnerable, actions by incumbents unwilling or incapable of heeding public opinion, shifting electoral coalitions that make extant majorities difficult to sustain, or entrenched ideological divisions in a district.

that unpopular decisions by elected officials will tempt challengers to enter a race, lowering an incumbent's vote share (cf. Brady, Canes-Wrone, and Cogan 2002). However, incumbents who make unpopular decisions may perceive themselves to be safe from effective challenges for reasons the analyst cannot observe. Similarly, one could argue that low levels of voter knowledge ought to be associated with greater incumbent noncompliance because voter ignorance allows incumbents to make choices without electoral implications. However, low levels of voter knowledge or attentiveness may be a consequence of incumbent compliance with voter demands and its deterrent effect on challenger entry, rather than a cause of noncompliance.

A more promising direction is to examine the behavior of incumbents. In particular, incumbents should alter their behavior given the threat of real electoral competition more than they would in its absence. Moving toward an empirical test of this prediction requires surmounting two challenges. First, we require a measure of the extent of incumbent compliance with public demands. Second, to overcome the endogeneity of challenger viability, we require an exogenous determinant of the potential threat of electoral competition.

2.3   The Case of Judicial Elections

Examining the behavior of trial judges is an especially useful means for understanding the connection between electoral institutions and the behavior of public officials. Examining the behavior of Kansas trial court judges in particular provides a unique opportunity to understand the specific effects of electoral competitiveness.

The behavior of elected trial judges. Trial judges must face periodic review by voters in 39 American states. Examining their behavior, as opposed to that of other elected officials, confers three immediate advantages. First, we can employ the sentences sanctioned by judges in individual criminal cases to obtain a direct measure of action that is comparable across officials and different institutional contexts. Owing to reporting requirements, data on sentencing is often accompanied by an enormous volume of information about specific conditions under which sentences were handed down. This makes it possible to account for much of the natural heterogeneity of circumstances judges confront when making decisions.

Second, voters tend to know next to nothing about judges' decisions in the vast majority of cases (Mathias 1990; Sheldon and Lovrich 1983). To the extent that very low levels of voter knowledge enhance opportunities for incumbent autonomy, the study of elected judges comprises a difficult case for testing general theories of electoral accountability.

Third, on the rare occasion that voters do become aware of judicial behavior, it is usually due to coverage of high-profile trials or controversial cases of recidivism. Critically, adverse publicity nearly always corresponds to cases of perceived judicial *leniency*. Media accounts of courtroom proceedings tend to result in voters believing judges are too lenient (Roberts and Edwards 1989). Additionally, voters are inclined to believe the criminal justice system as a whole is too lenient (Warr 1995). Finally, nearly all convicts claim to have been punished too much, and more definitive evidence of overpunishment typically comes to light years after a judge hands down a sentence.[3] By contrast, an episode of recidivism or unusually pointed criticism of a judge by a victims' rights group or police union (or challenger) provides a more immediate signal that a judge's sentence did

---

[3]Moreover, even evidence of wrongful prosecution or juror error attaches not to the sentencing judge, but rather to those who fabricated evidence or misjudged its veracity.

not fit the crime.

An informational environment in which judges have greater reason to fear voters perceiving them as too lenient than too severe (if they perceive judges at all) creates an asymmetry: If the constraint of public opinion binds at all, it will tend to make judges weakly more *punitive* rather than more moderate with respect to constituent preferences (Huber and Gordon 2004). The (weak) unidirectionality of the electoral incentive is valuable because it eliminates the need for the analyst to determine which direction (more lenient or harsh) engenders greater compliance with public demands for a particular judge, a difficulty that besets the analysis of legislative behavior (but see Lee, Moretti, and Butler 2004). This is critical because in general we will not be able to obtain a measure of a judge's primitive sentencing preferences apart from those induced by his or her political environment.

Kansas judicial selection as a natural experiment. There are 31 judicial districts in Kansas, each composed of from one to seven of the state's 105 counties. Kansas is one of two states in which the method of selecting judges differs from district to district. The state constitution holds that judges will be selected via partisan competitive election. Since 1972, however, subject to approval in a district-wide referendum, individual districts can choose to replace the default mechanism with a system of gubernatorial appointment and non-competitive retention elections. Incumbent judges occupy unique seats, and are therefore not in direct competition with one another. Judges in all districts serve staggered four year terms, with elections occurring in even years.

A map of the judicial districts appears in Figure 1. Fourteen districts employ the familiar partisan selection method. These districts comprise 53 counties, in which roughly 42% of Kansans reside. Judges seek their party's nomination in primaries held in August.[4] The victors then face off in a general contest on Election Day in November.[5] Seventeen districts employ the alternative method. A nominating commission in each proposes names to the governor.[6] If she chooses to appoint a nominee, that judge will serve a one-year probationary term followed by a retention vote. If retained, she will stand for retention every four years thereafter. In a retention election, the judge's name appears on the ballot, and voters can vote "Yes" or "No."[7]

Assuming that a district's choice of method for selecting judges is unconfounded by factors that influence judicial decision making (discussed in detail below), the institutional variation in Kansas judicial districts permits us to consider the incentives for incumbents created by the threat of challengers while circumventing the problem of endogenous challenger viability. Whereas the threat posed by challengers in the primary or general election may vary among competitive districts,

---

[4]The partisan primaries are closed to members of the opposite party, although the Democratic primary is open to unaffiliated voters.

[5]We compiled data on the 138 contests in partisan competitive districts involving an incumbent trial judge from 1994 to 2002 (source: Kansas Secretary of State). Nine incumbents faced challengers in primary elections (three lost) and ten did so in general elections (one lost). As we noted above, because challenger entry is endogenous to incumbent performance, these numbers by themselves do not indicate an effect on incumbent accountability, although they do reveal that incumbent judges can be defeated.

[6]Commissions are each composed of an equal number of attorneys and no attorneys. Attorney members are elected by their peers; others are appointed by Boards of County Commissioners.

[7]From 1994 to 2002, 144 judges stood for retention election; none lost his or her seat. The average Yes vote for incumbent judges was 76% (with a standard deviation of 5%) and the minimum was 51%.

Figure 1: Kansas Judicial Districts and Selection Rules
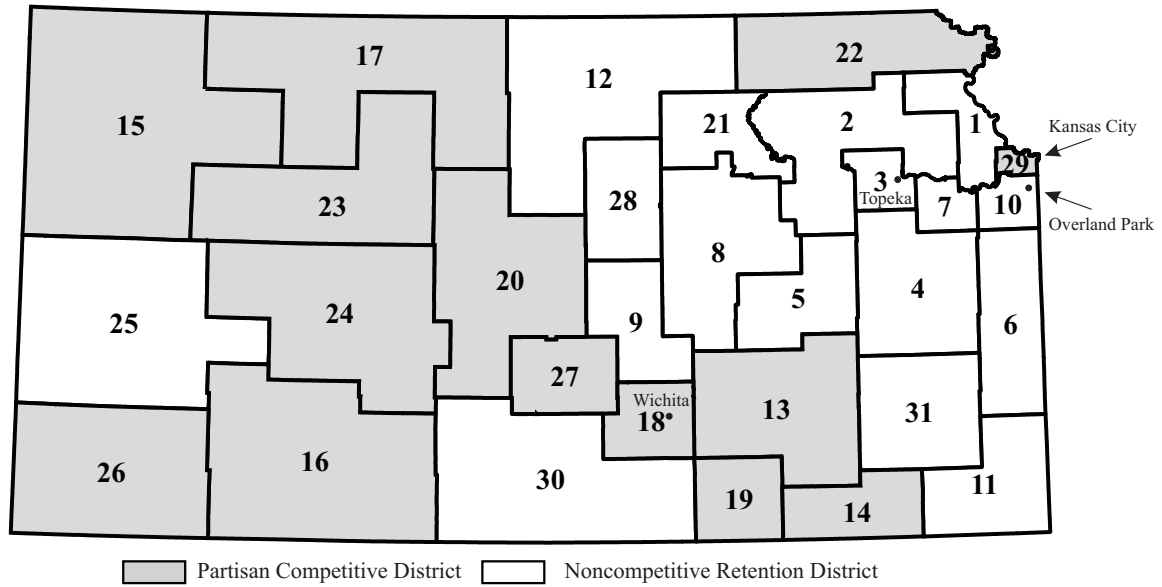
in retention districts there can be no challenger *irrespective of incumbent behavior in office.* The set of decisions made by judges in retention districts therefore constitutes a suitable control group against which to make comparisons of decisions made by judges who serve under the threat of primary and/or general election challenges.

Other empirical contexts in which these comparisons might be made are problematic. One could, for example, consider comparing judicial sentencing behavior across states with different selection methods. Even if one could adequately control for the contextual and institutional heterogeneity across states, however, fundamental differences in legal systems would remain difficult to account for. Criminal codes vary enormously in how they categorize crimes, and judges in different state have vastly different discretion in punishing offenders. By confining our analysis to a single state, we can hold constant the legal system under which judges (as well as prosecutors, defense attorneys, and defendants) operate.

Another possibility is to examine the behavior of officials in another state. For example, Missouri has a similarly bifurcated system of selecting judges. However, Missouri adopted nonpartisan selection of circuit court judges only in urban areas (Kansas City and St. Louis) on the heels of charges that urban political machines were exercising undue influence in the selection process (Watson and Downing 1969). The effect of the selection mechanism in Missouri is therefore not separable from numerous other differences between urban and rural counties. Such confounding influences are likely to be minimal in the Kansas setting. Two of the four most urban counties in the state (Wyandotte and Sedgwick) select district judges via partisan races, while the other two (Shawnee and Johnson) employ a retention system. Rural counties are similarly split.

To determine whether the institutional variable is a proxy for other features of judges' environments, we gathered data on the political and demographic characteristics of Kansas' 31 judicial districts. We then compared the characteristics of the partisan competitive and retention districts using t-tests of equality of means and bootstrapped Kolmogorov-Smirnov tests of equality of distri-

Table 1: Comparing District Characteristics in Partisan Competitive and Retention Districts: t and Bootstrapped Kolmogorov-Smirnov Balance Tests

	(1) Before District Matching		(2) After District Matching	
	t-test p-value	KS-test p-value	t-test p-value	KS-test p-value
Democratic presidential vote share[1]	0.34	0.25	0.91	0.84
Proportion with at least some college[2]	0.44	0.15	0.30	0.28
Proportion black[2]	0.64	0.72	0.49	0.53
Propotion nonwhite[2]	0.58	0.61	0.23	0.83
Population Density[2]	0.93	0.10	0.26	0.52
Proportion urban[2]	0.63	0.63	0.85	0.83
Income[2]	0.14	0.13	0.48	0.51
Crime rate[3]	0.02	0.08	0.32	0.85

[1] 2000, Kansas Secretary of State. [2] 2000, Bureau of the Census. (Note: For districts with multiple counties, income is caculated as the population-weighted average of county-level median household income.) [3] 2002, FBI Uniform Crime Statistics.

butions. The first two columns of Table 1 report the associated p-values.[8] District crime rate is the only variable for which substantial imbalance exists between the two systems – retention districts have significantly higher crime rates, a result largely driven by Shawnee County (the location of Topeka).

2.4   Identification Strategy

In Appendix A, we present a simple heuristic model to clarify the intuition underlying the incentive effect of potential competitiveness on elected judges. We anticipate that, ceteris paribus, incumbent judges in Kansas's partisan competitive districts will sentence more punitively than judges in its retention districts. While consistent with an incentive effect, such a finding would also be consistent with a selection account: voters in partisan districts may select inherently more punitive judges than judicial nominating commissions do in retention districts. Alternatively, voters in partisan districts may elect district attorneys more punitive than voters do in retention districts.

Addressing the concern about differences in elected prosecutors is straightforward and is discussed in greater detail below. Our strategy for identifying the incentive effect of potential competitiveness, and for eliminating rival interpretations for that effect, proceeds in three phases. The first two aim to ensure that the distinction between electoral systems is not merely a proxy for other features of districts that may contribute to differences in sentencing. Phase one is regression-based: we consider whether the estimated effect of the electoral system is sensitive to the inclusion of observable district characteristics, including the crime rate, in a saturated model of sentencing behavior. Phase two relaxes the parametric assumptions of regression in a matching analysis. We compare outcomes from similar criminal cases drawn from matched pairs of observably similar judicial districts with different electoral systems. This approach discards data from retention districts that are not comparable to those that still employ the competitive system (e.g. Shawnee County).

[8]A probit model predicting selection system yielded substantively identical results.

Of course, the possibility still remains that the incentive effect of the electoral system is confounded by an unobservable tendency of partisan competitive districts to select inherently more punitive judges. Our strategy for disentangling these competing explanations for differences in judicial behavior (Phase three) exploits variation in judicial behavior over their electoral calendars. If elected judges discount the future value of holding office relative to the benefit of assigning their most preferred sentences, then an incentive-based account predicts that judges will sentence weakly more punitively as election approaches, controlling for secular trends in sentencing. We refer to this phenomenon as the "proximity effect." A pure selection account, by contrast, is incompatible with the dynamic adjustment implied by a significant proximity effect.

More importantly, observing the nature of the *interaction* between electoral rules and electoral proximity sheds additional light on how judges' incentives operate. In the model in Appendix A, we demonstrate that incumbent response to the threat of a potentially viable challenger is consistent with either a larger *or* smaller proximity effect in partisan competitive districts than in retention districts. We anticipate a larger effect in the partisan districts if, for example, judges strongly discount the future value of office or place substantial weight on assigning their most preferred sentences. Likewise, we anticipate a larger proximity effect in the retention districts if judges discount the future value of holding office only slightly or place relatively little weight on their own sentencing preferences.

It should be noted that detecting a larger proximity effect in retention than in partisan districts, while consistent with the role we ascribe to challengers, would also be consistent with the following alternative account: Suppose nominating commissions in retention districts select judges systematically more lenient than those in partisan districts. Concerned with negative coverage from the media and interest groups, inherently lenient judges in retention systems behave more punitively as voter review approaches. In contrast, their counterparts in the partisan competitive districts do not display a proximity effect because those judges are punitive to begin with. In this account, the observed effects (judges in partisan districts more punitive on average; proximity effect in retention districts but not partisan districts) would be generated independently of the threat a potential challenger poses.

To summarize: (1) A finding that at least some judges become more punitive as election approaches is consistent with an account based on electoral incentives and not selection alone; (2) A finding that this electoral proximity effect is greater in retention districts than partisan competitive districts is consistent with both the challenger-based incentive mechanism and the alternative account; (3) A finding that the proximity effect is greater in partisan competitive districts than retention districts is consistent with the challenger-based incentive mechanism but *not* the alternative account.

3   Data and Method

3.1   Sentencing in Kansas

We obtained data on the sentencing behavior of 160 Kansas district court judges from 1997 to 2003. Seventy-three judges served in districts with partisan competitive elections, and 87 in districts

with retention elections.[9] Criminal sentencing in Kansas is governed by the Kansas Sentencing Guidelines Act of 1993 as amended (hereafter, the guidelines), which places limits on the discretion of trial court judges in assigning sentences to convicted offenders. Convicts serve at least 85% of the sentence a judge hands down before becoming eligible for parole. As with most guideline systems, judges are required to take into account an offender's criminal history and the severity of the offense committed to determine an applicable range of appropriate sentences. Judges then have limited discretion to depart from the recommended range.

Although the guidelines suggest a single, presumptive sentence for history/severity combinations, judges can generally choose to assign any sentence between the minimum and maximum guideline sentences without further justification. If a judge wishes to "upwardly depart," he or she may assign a sentence up to twice the guideline maximum given one or more aggravating circumstances specified in the statute.[10] He or she may also assign a departure sentence below the minimum sentence given mitigating circumstances.[11] Departure sentences are subject to appellate review and will be sustained if there are "substantial and compelling reasons for the departure."[12] In multiple count convictions, judges have discretion to assign sentences on less severe counts to run either concurrently or consecutively to the first sentence with the condition that the total time in prison cannot exceed twice the sentence on the primary count.

The dataset for our analysis was created by merging information about Kansas's judges with records of the sentences they assigned to individual defendants collected by the Kansas Sentencing Commission. We restricted our analysis to those felonies for which there were a reasonable number of cases across the state (more than 250), for which judges have discretion in sentencing, and for which incarceration is a possibility. This left us with a range of person (assault, criminal threat, robbery, sexual assault) and property crimes (theft, burglary, arson).[13] We have 18,139 cases for the period between July 1, 1997 and June 30, 2003.[14]

---

[9]Some districts also have magistrate judges, who have limited authority and whose decisions are excluded from our analysis.

[10]Prior to the Kansas Supreme Court's decision on May 25, 2001 in *State v. Gould* (271 Kan. 394, 23 P.3d 801), judges were free to identify those aggravating factors that warranted a departure above the guideline maximum sentence. The *Gould* decision held that any facts that led a judge to assign a sentence above the guideline maximum would have to be proven "beyond a reasonable doubt" before a jury. In a measure that became effective June 6, 2002, the state legislature altered the guidelines to require such factors be proven to a jury, either during trial or in a separate post-conviction sentencing hearing. In our analysis, we account for how this ruling, in the period between May 2001 and June 2002, limited judges' authority to sentence single count cases to no more than the guideline maximum and multi-count cases to no more than twice that quantity.

[11]The *Gould* decision did not affect these downward departures.

[12]See K.S.A. 22-3604.

[13]We excluded homicide cases for two reasons. First, under Kansas law, judges lack sentencing discretion for murder convictions. (The jury decides between life imprisonment and the death penalty.) Second, defendants often plead guilty to manslaughter to avoid a murder trial. (Including the manslaughter cases does not affect our main results.) Additionally, we excluded drug crimes on the grounds that preferences about punishment for drug offenders may vary substantially, whereas punishment for those convicted of the crimes we examine is uncontroversial.

[14]We discard all cases heard by judges who sentenced fewer than 25 cases in our dataset and all cases heard by retention judges during their probationary term. Neither restriction affects our results.

The vast majority of cases were resolved via plea bargain. This is similar to the situation in most states, and does not threaten our ability to make inferences about judges' incentives. A judge's preferred sentence (discounted by the probability of a conviction) is properly viewed as a reversion or threat point in the negotiation between the prosecutor and defendant. Also, judges in Kansas have the discretion to reject settlements between prosecutors and defendants.[15] Because bargaining takes place in the shadow of the judge (LaCasse and Payne 1999), it incorporates the judge's underlying preferences about appropriate punishment.[16] Observed sentences range from zero to 3,237 months. In 69% of cases, the sentence includes probation, a fine, or community service, but no time in prison. For the remaining 31% of cases, the median prison sentence is 32 months. The number of counts in a conviction ranges from one to 50. 74% of cases have only a single count, and 99% have 5 or fewer counts.

3.2 Descriptive Analysis

A cursory examination of the data reveals preliminary support for the hypothesis that sentencing in partisan competitive districts is more severe than in retention districts. 35% of sentences handed down in the competitive districts include prison terms, compared with 27% in the retention jurisdictions. Likewise, the median non-zero prison sentence is higher in the partisan than retention districts – 33 versus 31 months. (The respective average non-zero prison terms are 66 and 57 months. All of these differences are highly statistically significant).

Suppose, however, that the distribution of sentences differed in its higher-order moments as well. If, for example, the variance of sentences was higher in competitive systems because judicial "temperament" differed systematically, then a risk-acceptant defendant might prefer the gamble of being sentenced by a judge in a partisan competitive district. We might then conclude that judges in retention systems were, in some sense, more punitive. To confront this possibility, we conduct a graphical test of stochastic dominance. Consider a defendant who could choose whether to be sentenced by a judge serving in a competitive district or a retention district. Let $F_r(\cdot)$ be the cumulative distribution function (cdf) of sentences in retention systems, and $F_c(\cdot)$ be the corresponding cdf in competitive districts. The competitive distribution first-order stochastically dominates the retention distribution if $F_c(x) < F_r(x)$ for all x. In the context of sentencing, where defendants presumably want the *smallest* possible sentence, stochastic dominance of sentencing in competitive districts would imply that any rational individual, irrespective of his or her risk preferences, would prefer being sentenced in a retention district.

We divided the sentences by electoral system and graphed their quantile-quantile (QQ) plot. This QQ-plot, displayed in the top-left panel of Figure 2, shows the percentiles of each distribution graphed against each other, along with a 45° degree line for comparison. (The hash marks on the horizontal axis suggest the location of the preponderance of the data.) Dominance of one distribution versus another is implied if the plot lies always above or always below the diagonal. The plot in Figure 2 is consistently above the diagonal over its entire range except at the highest quantile. An examination of the underlying data reveals that the point at which the QQ-plot does finally cross the diagonal corresponds to 5 cases out of 18,139, or three hundredths of one percent

---

[15]Note that the fact that judges rarely reject plea bargains is not evidence against a judge's influence in the process; if attorneys correctly anticipate what a judge will accept, we should never observe plea bargains rejected.

[16]Confirming the judge's importance, we find substantial variation in sentencing practices in districts with multiple judges but only one district attorney.

Figure 2: Quantile-Quantile Plots of Sentencing Distributions by Selection Mechanism
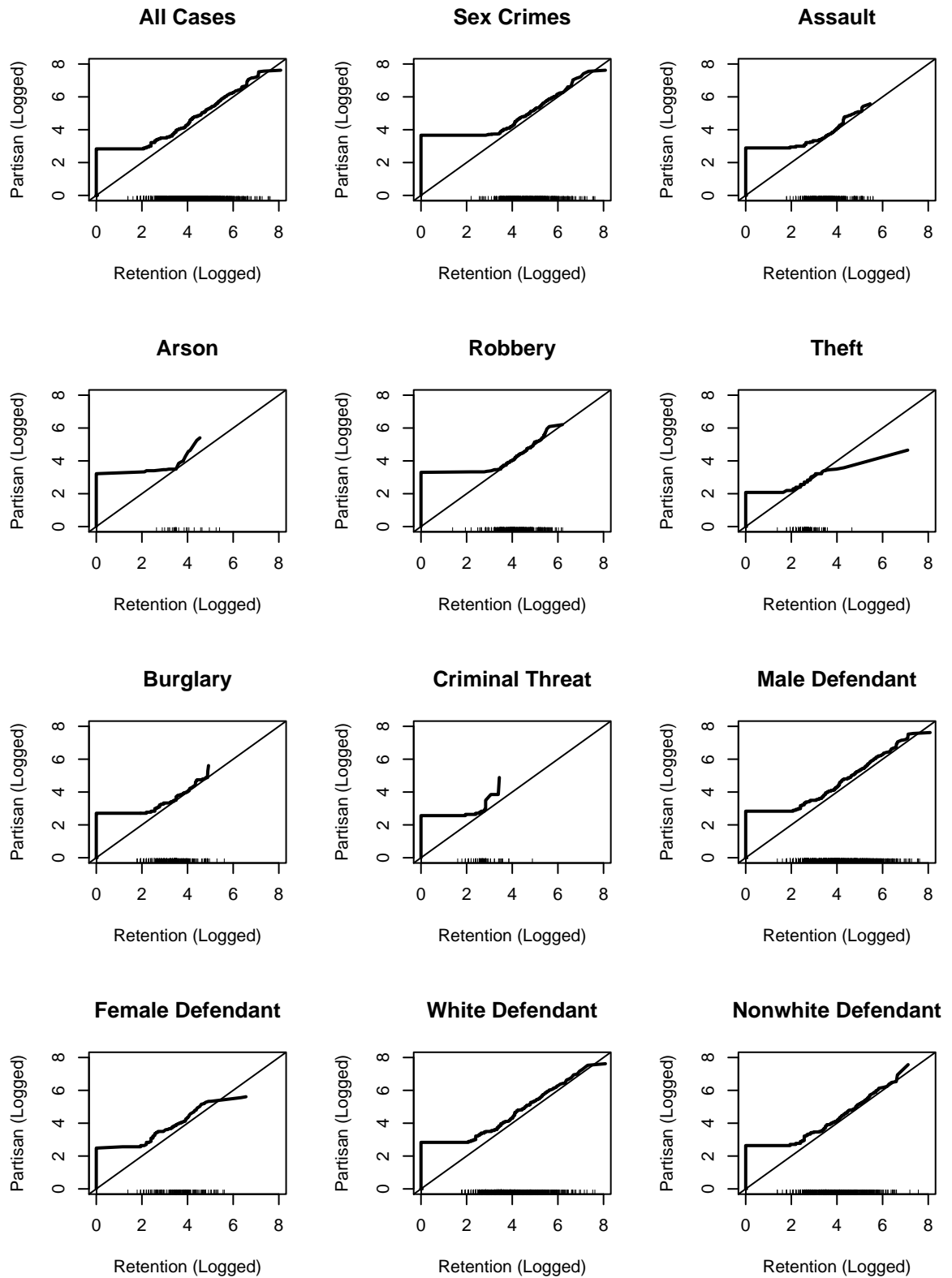
of the sample. Each of these was a sex crime occurring in Sedgwick County (Wichita). Ignoring these five cases, we would conclude that any random defendant would prefer being sentenced by a judge in a retention system to one in a competitive one, irrespective of his or her risk preferences.

The result does not appear simply to be a consequence of aggregating over different crimes and categories of defendants. Figure 2 also displays QQ-plots disaggregated by crime, sex, and race. With one exception, the plots consistently remain above the diagonal line, suggesting dominance of sentencing in partisan competitive districts. The QQ-plot for theft deserve special attention. For that crime, the portion of the plot falling below the diagonal corresponds to only 22 cases, about one-half of one percent of all thefts.

3.3   Case-level Covariates

A downside of the above approaches is that they do not control for other relevant distinctions among cases. Fortunately, we have numerous variables to account for contextual heterogeneity in sentencing. Summary statistics for model variables appear in Table 2. Our primary measure of culpability is the natural log of the presumptive sentence (plus one, to match the scaling of the dependent variable) associated with the conviction's top count. This captures the extent to which the state's elected officials view that criminal act by a defendant with a particular criminal history as harmful. We also include two indicator variables to control for revisions to the guidelines that occurred in 1996 and 1999.[17] Because judges have discretion to sentence additional counts concurrently or consecutively (subject to constraints discussed above), we also control for the number of additional counts in the conviction. Aggravating factors the judge is obliged to take into account include whether the defendant is a classified persistent sex offender, whether he or she was in possession of a firearm at the time of the crime, and whether the victim was a child or a government employee or law enforcement official.

We also include crime-specific indicator variables to account for heterogeneity in sentences, and year-specific effects to capture non-monotone secular trends in perceptions of criminal culpability. These latter controls are also vital because we wish to account for the confounding effect of the *prosecutor's* electoral calendar. All district attorneys in Kansas are on the same four year electoral calendar, but judges within a district serve staggered terms. Year effects therefore control for changes in sentencing that might result from variation over time in prosecutors' electoral incentives. In our analysis of the judge's electoral proximity effect, we can also account for average differences in district attorneys' offices through the inclusion of judge-specific fixed effects. These span all time-invariant features of the district (including, for example, the culture in the prosecutor's office).

Variables affecting punitiveness consist of defendant-, case-, and judge-specific characteristics. Included in the model are indicators of whether the sentence resulted from a plea bargain and whether the defendant had appointed counsel. We also include measures representing whether the defendant was male, nonwhite, or Hispanic. We control for defendant age and age-squared (under the hypothesis that judges are likely to be lenient toward both the youngest and oldest defendants). An indicator variable equals one if the judge serves in a partisan competitive district, and zero if in a retention district. Finally, our measure of electoral proximity ranges from zero to one, with one implying imminent electoral threat. A comprehensive discussion of how the measure is constructed appears in Appendix B.

---

[17]Earlier guideline regimes are relevant because the applicable guidelines are those in place at the time of the felony, not the sentencing.

Table 2: Summary Statistics for Model Variables

Variable	Mean	Std. Dev.	Min.	Max.
Ln(1+assigned prison sentence in months)	1.12	1.74	0	8.08
Ln(1+presumptive prison time in months)	1.02	1.76	0	6.66
First guideline regime (1=yes)	0.03	0.16	0	1
Second guideline regime (1=yes)	0.42	0.49	0	1
Additional counts in conviction	0.46	1.41	0	49
Persistent sex offender (1=yes)	0	0.06	0	1
Firearm used in commission of crime (1=yes)	0.03	0.18	0	1
Victim government official (1=yes)	0.02	0.14	0	1
Victim child (1=yes)	0.09	0.28	0	1
Assault (1=yes)	0.21	0.41	0	1
Criminal Threat (1=yes)	0.07	0.26	0	1
Robbery (1=yes)	0.08	0.27	0	1
Sex Offense (1=yes)	0.11	0.31	0	1
Theft (1=yes)	0.23	0.42	0	1
Burglary (1=yes)	0.28	0.45	0	1
Appointed counsel (1=yes)	0.78	0.41	0	1
Plea bargain (1=yes)	0.94	0.24	0	1
Defendant male (1=yes)	0.89	0.31	0	1
Defendant nonwhite (1=yes)	0.28	0.45	0	1
Defendant Hispanic (1=yes)	0.08	0.28	0	1
Defendant age (years)	29.26	10.02	15.44	86.1
District Democratic voteshare	0.39	0.11	0.2	0.67
District crime rate	24.62	19.23	3.91	73.45
District proportion nonwhite	0.17	0.12	0.02	0.42
District population density	398.99	398.33	4.04	1045.58
District proportion urban	0.75	0.22	0.13	0.95
Partisan (competitive) district	0.50	0.50	0	1
Electoral Proximity	0.55	0.28	0	1

N=18,139. Sentencing data from Kansas Sentencing Commission. Elections data from Kansas Secretary of State. See Table 1 for sources for district demographic data.

4 Results

4.1 The Baseline Systemic Effect

Our first set of estimates concerns the baseline hypothesis: Ceteris paribus, judges should behave more punitively in partisan competitive districts than retention districts. We first present the results of our regression estimation, and then proceed to the matching analysis.

Regression Analysis. Our regression analysis relies on two substantive assumptions. First, all judges, no matter how lenient, adhere to the principle of *proportionality*: greater culpability demands greater punishment. Second, from the perspective of the defendant, the worst non-incarcerative sentence is preferable to the most lenient prison term. In Appendix C, we demonstrate how these assumptions, coupled with the right-censoring implied by statutory maximum sentences, yields the two-limit Tobit likelihood function (with an adjustment for judge-specific groupwise heteroscedasticity). The dependent variable is the natural logarithm of the prison term, in months, plus an unidentified constant. We adopt the common practice of setting the constant to one. This normalization implies that one month in prison is exactly twice as painful as the worst non-incarcerative punishment. (We experimented with a broad range of alternative values, none of which affected our results.) Finally, note that a left-censored observation corresponding to zero prison time does not stand in for an unobserved "negative" sentence. All convicted felons are punished, but many are punished with a sentence less severe than prison.

We estimated four Tobit models, the coefficient estimates from which appear in Table 3. Specifications (1) and (2) employ all observations. Specifications (3) and (4) restrict the sample to all cases from retention districts plus cases from competitive districts in which the judge faced no competition in the subsequent election. Restricting the sample in this way is meant to capture the notion that it is the *potential* for challenger entry, rather than the fact of a challenge, that motivates changes in the behavior of incumbents. Models (2) and (4) include aggregate district-level measures in addition to case- and defendant-level factors.

We note a large number of significant predictors of punishment. The presumptive sentence and additional counts on the conviction have very strong positive effects, as expected, as do the aggravating factors. Also as expected, the presence of appointed (rather than privately-hired) counsel raises the expected length of incarceration while a plea bargain lowers it. We further find that nonwhites and Hispanics tend to receive larger punishments, even controlling for the legally relevant characteristics of individual cases. These findings, while troubling, are beyond the scope of the current analysis.

The column (2) and (4) specifications with district-level measures suggests more stringent sentences in districts with larger nonwhite populations and with higher crime rates. Controlling for population density, it is more rural districts that have more stringent sentences. Finally, likelihood ratio tests for all specifications allow us overwhelmingly to reject the null hypothesis that the year-specific indicator variables are jointly insignificant.

Next, we turn to the effect of potential electoral competition. In all specifications, as expected, the coefficient on the partisan competitive district indicator is positive and highly statistically significant. Note that restricting our sample to judges who were subsequently unchallenged does not substantially alter the estimated effect of competitiveness. This suggests that the results are not driven by the behavior of judges who, ex ante, perceive themselves as particularly vulnerable.

13

Table 3: The Effect of Judicial Selection Mechanism on Sentencing (Tobit estimates)

	All Judges		Unchallenged Judges	
	(1)	(2)	(3)	(4)
Ln(1+presumptive prison time in months)	1.14	1.15	1.15	1.16
	(49.42)	(49.57)	(48.58)	(49.48)
First guideline regime	0.37	0.30	0.41	0.34
	(2.18)	(1.85)	(2.32)	(1.99)
Second guideline regime	0.25	0.22	0.23	0.20
	(2.26)	(2.05)	(1.96)	(1.67)
Additional counts in conviction	0.14	0.15	0.15	0.16
	(7.27)	(7.45)	(6.99)	(7.18)
Persistent sex offender	1.87	1.91	1.95	2.00
	(4.60)	(4.75)	(4.13)	(4.28)
Firearm	1.56	1.59	1.60	1.64
	(10.81)	(11.07)	(10.00)	(10.25)
Victim government official	0.95	0.95	0.86	0.86
	(5.04)	(4.85)	(4.17)	(4.01)
Victim child	-0.12	-0.10	-0.10	-0.10
	(0.74)	(0.66)	(0.62)	(0.58)
Appointed counsel	0.92	0.90	0.96	0.94
	(11.90)	(12.11)	(11.56)	(11.81)
Plea bargain	-0.99	-1.03	-1.02	-1.05
	(10.64)	(10.92)	(10.64)	(10.79)
Defendant male	1.31	1.29	1.33	1.32
	(10.53)	(10.64)	(10.01)	(10.11)
Defendant nonwhite	0.11	0.11	0.11	0.09
	(1.80)	(1.76)	(1.64)	(1.44)
Defendant Hispanic	0.25	0.14	0.30	0.19
	(2.37)	(1.39)	(2.61)	(1.74)
Defendant age (years)	0.14	0.14	0.14	0.14
	(9.05)	(9.00)	(9.49)	(9.39)
Defendant age squared	-0.00	-0.00	-0.00	-0.00
	(7.61)	(7.58)	(7.97)	(7.91)
District Democratic voteshare		-0.88		-1.06
		(1.25)		(1.54)
District crime rate		0.01		0.01
		(3.27)		(3.81)
District proportion nonwhite		1.79		2.09
		(2.51)		(2.82)
District population density		0.00		0.00
		(2.41)		(2.80)
District proportion urban		-1.18		-1.32
		(3.23)		(3.87)
Partisan (competitive) district	0.26	0.42	0.26	0.40
	(2.39)	(3.01)	(2.49)	(2.71)
σ	2.63	2.62	2.66	2.64
log-likelihood	-18070.9	-18036.29	-16319.28	-16282.92

N=18,139 in columns (1) and (2) and 16,526 in columns (3) and (4).
Dependent variable is Ln(1+assigned prison time in months).
Absolute value of groupwise (judge) heteroscedasticity-consistent t-statistics appear in parentheses.
Coefficients for Crime and Year dummies are not reported.

Because the magnitude of Tobit coefficients can be difficult to interpret, we derived several quantities of greater substantive interest. This entailed setting all control variables at their sample medians (for district characteristics, we employ the district medians), and employing the modal crime (burglary) and year (2001). Our findings suggest estimated differences in the probability of an assigned prison term between partisan competitive and retention districts from 2.7% and 4.3%, depending on specification. These numbers may seem small at first, but one must keep in mind that the baseline probability of incarceration with the control variables set in the way described above is about 18%. The proportionate increase in the probability of incarceration associated with a change in the electoral rules is therefore about 15% to 25%, depending on specification.

Next, we calculated the change in the expected sentence given prison assignment (a more meaningful quantity than the change in unconditional expected prison time). Here, we set the value of the logged presumptive sentence equal to its median among observations for which prison was imposed. We find that the presence of a potential challenger increases the expected non-zero sentence by about 2.46 to 3.83 months, which represents a 7.5% to 12% increase over the median non-zero prison sentence in a retention district (32 months).

Matching. The Tobit models, while theoretically supported, require strong functional form assumptions (e.g. linearity in variables). To ensure these do not drive our results, we also analyzed the data using a more flexible approach: nearest neighbor, one-to-one matching.[18] Matching proceeds by pairing observations from treatment (partisan competitive) and control (retention) groups that are similar in terms of their observed covariates, and comparing the outcomes (incarceration). In the current application, the wealth of data permit us to obtain *exact matches* – and thus perfect balance – in treatment and control groups on all discrete case, defendant, and crime characteristics listed above. Given a group of exact matches, we pair observations closest in defendant age.[19]

Estimates for the matching analysis are displayed in Table 4. The table displays estimates of average treatment effects on the treated (ATT), i.e. the estimated effect of the electoral system on sentences administered in partisan competitive districts.[20] For the ATT estimates in columns (1) and (2), we make no effort to achieve balance on district-level observables. A case with a particular fact pattern from a partisan competitive district may be matched with one with the same fact pattern from any of the seventeen retention districts. The first row of estimates reports ATT estimates of changes in the probability of incarceration. The highly statistically significant 3.4 to 3.6% increase confirms the Tobit results.

The parametric assumptions underlying the Tobit specification permitted us to calculate the expected change in incarceration given a prison sentence was imposed. The matching algorithm does not permit us to derive a comparable figure without additional assumptions; however, by pooling both prison and non-prison sentences, we can estimate the average treatment effect in months across all observations. The expected shift – quite small because of the large number of sentences with no prison time – is a still statistically significant 0.48 or 0.51 months, depending on whether subsequently challenged judges are included.

Our ability to exactly match on discrete case factors stems from the fact that there are 1,134 unique fact pattern "clusters" in the sentencing data for which we possess observations from both

---

[18]Matching was conducted using Sekhon's (2005) *Match* algorithm in *R*.

[19]When necessary, we employ a caliper to insure balance on age. The *p*-values for bootstrapped Kolmogorov-Smirnov tests of equality of distributions after matching range from 0.19 to 0.92.

[20]ATT estimates permit superior balance on covariates not exactly matched on (particularly at the district level – see below), but are comparable to average treatment effect (ATE) estimates.

Table 4: Matching Estimates of the Average Effect of Treatment (Potential Competition) on the Treated (Cases in Partisan Competitive Districts)

| | Fact Pattern Matching Only | | District, Then Fact-Pattern Matching | |
| | (1) | (2) | (3) | (4) |
Expected Change in . . .	All Judges	Unchallenged	All Judges	Unchallenged
Pr(prison)	0.034	0.036	0.067	0.08
	(5.64)	(5.57)	(18.17)	(19.99)
Months prison (unconditional)	0.513	0.48	1.15	1.406
	(3.43)	(2.91)	(2.91)	(2.91)
Matched treated observations	4,717	3,868	1,195	1,016

treatment and control groups. Distance matching on defendant age within each cluster without concern of ties is expedited by the presence in the data of defendant birthdays, which allow us to measure this characteristic to the day. Matching observations on the basis of district-level variables is more complicated, owing to the presence of just 31 unique values for each measure. This essentially guarantees the district-level covariates will not be balanced between treatment and control groups at the case level, even if excellent balance can be achieved at the district level. We therefore adopt a two-step approach. First, using the genetic matching algorithm of Diamond and Sekhon (2005), we pair partisan competitive districts with politically and demographically similar retention districts. For example, District 18 (Sedgwick County, the location of Wichita) is paired with District 7 (Douglas County, the location of Lawrence). (The full list of district matches is: 13/9, 14/8, 15/12, 16/8, 17/12, 18/7, 19/21, 20/9, 22/2, 23/31, 24/12, 26/25, 27/9, and 29/7.) Post-matching balance statistics at the district level appear in the third and fourth columns of Table 1. Note that this technique enables us to achieve excellent balance on the crime rate, which was significantly unbalanced in the raw district-level data. Having matched comparable districts, we then search for unique fact pattern clusters in district pairs, matching observations closest in age as above.

Results from the analysis with district matching appear in the third and fourth columns of Table 4. The district matching technique discards a large volume of sentencing information, drawing cases from only eight of the seventeen retention districts to assure comparability. For example, only one of the five southeastern retention districts districts (31) is kept. Shawnee County, with its unusually high crime rate, is discarded.[21] This approach dramatically decreases the overall sample size compared to matching on case-level covariates only. However, as the table indicates, adopting the more conservative approach approximately *doubles* the magnitude of the estimated effects, which remain highly statistically significant. This again confirms our basic result: cases with observably identical fact patterns and defendant characteristics are more likely to result in stiffer penalties in districts where the threat of electoral competition looms over the judge.

---

[21]Unfortunately, we lack historical data on the vote margins with which different districts adopted the noncompetitive retention method. Shawnee County is an exception, however. In 2000, residents voted overwhelmingly, 62-38%, to keep the retention system, which it initially adopted in 1974. This result lends face validity to the district matching technique, which drops cases from Shawnee altogether.

4.2 Exploiting Electoral Proximity to Evaluate Competing Mechanisms

The foregoing analysis has provided strong evidence that judges in partisan competitive districts are more punitive than those in retention districts. This finding is consistent with incumbent incentives generated by the potential informational role of challengers, but also, as discussed above, with alternative causal mechanisms. As we note in Section II, however, a finding that judges in partisan competitive systems become more punitive as election draws closer while judges in retention systems do not would constitute empirical confirmation of the challenger-induced incentive account, and disconfirmation of the alternatives.

Table 5 provides empirical support for this prediction. We created interaction terms to estimate the effect of electoral proximity in each type of system. (This is equivalent to, although easier to interpret than, including the baseline effect of electoral proximity across districts, the partisan competitive indicator, and the interaction of electoral proximity and the partisan competitive indicator.) The specification in column (2) includes judge-specific fixed effects, which permit us to control for all time invariant characteristics of a judge's (and her district's) punitiveness – including those arising from differences among districts in the kinds of judges (and prosecutors) they tend to put on the bench. The fixed effects span the selection mechanism indicator variable and observable district-level characteristics, so those measures are omitted from the second specification.

The primary variables of interest are the effect of selection system in (1) and the interaction between the institution and electoral proximity in both (1) and (2). The coefficient estimates for these variables confirm three important points. First, electoral proximity exerts a statistically significant pressure to become more punitive in the partisan competitive districts, but *not* in the retention districts. Holding the control variables at the same values discussed above, a shift from minimal proximity (e.g. the day after a filing deadline in which no challengers have filed) to maximal proximity (election season in the presence of a challenger) in a competitive district leads to a statistically significant 3.1% increase in the probability of incarceration and, conditional on incarceration, a sentence three months longer. These effects are substantial, given that the baseline probability of incarceration is around 18% and the median sentence length in cases involving any incarceration is 32 months. Thus, these figures represents a 17% proportional increase in the probability a defendant is incarcerated and a 9.4% increase in sentence length. (The effects are even larger in the column [2] specification.) In noncompetitive districts, by contrast, an increase from minimal to maximal electoral threat produces a *decrease* in the probability of incarceration and sentence length, but in neither specification is that difference statistically distinguishable from zero.

Second, the coefficient on the first-order partisan competitive indicator in specification (1) is statistically indistinguishable from zero. This suggests that under conditions of minimal electoral threat (when proximity is zero), judges in partisan and retention districts *do not differ* from each other in their sentencing in a statistically meaningful way – evidence against the argument that selection, rather than incentives, drives baseline differences in sentencing between the two systems.

Finally, an *F*-test reveals that under conditions of maximal electoral threat (when electoral proximity is one), judges in partisan districts are significantly more punitive than judges in retention systems. We can also determine how far into a term typical judges are before the difference is significant. Simulation reveals that once a partisan judge is 33% into her term (about one and a third years), his or her sentencing becomes significantly more punitive than a similarly situated retention judge.

Table 5: The Effect of Electoral Proximity on Sentencing by Selection Mechanism (Tobit Estimates)

	(1)	(2)
Ln(1+presumptive prison time in months)	1.15	1.15
	(49.78)	(75.33)
First guideline regime	0.30	0.27
	(1.87)	(1.64)
Second guideline regime	0.22	0.19
	(2.02)	(1.91)
Additional counts in conviction	0.15	0.15
	(7.46)	(7.63)
Persistent sex offender	1.90	1.93
	(4.74)	(5.16)
Firearm	1.59	1.56
	(11.04)	(12.78)
Victim government official	0.94	0.93
	(4.85)	(5.78)
Victim child	-0.09	-0.14
	(0.6)	(0.96)
Appointed counsel	0.90	0.91
	(12.10)	(12.80)
Plea Bargain	-1.03	-0.97
	(10.97)	(11.85)
Defendant male	1.29	1.30
	(10.62)	(11.57)
Defendant nonwhite	0.11	0.08
	(1.74)	(1.29)
Defendant Hispanic	0.14	0.12
	(1.37)	(1.25)
Defendant age (years)	0.15	0.15
	(9.03)	(10.11)
Defendant age squared	-0.00	-0.00
	(7.62)	(8.35)
District Democratic voteshare	-0.88	
	(1.25)	
District crime rate	0.01	
	(3.29)	
District proportion nonwhite	1.80	
	(2.52)	
District population density	0.00	
	(2.43)	
District proportion urban	-1.18	
	(3.24)	
Partisan (competitive) district	0.14	
	(0.78)	
Partisan (competitive) district × Electoral Proximity	0.27	0.29
	(1.93)	(2.45)
Retention (noncompetitive) district × Electoral Proximity	-0.24	-0.24
	(1.47)	(1.71)
σ	2.62	2.57
log-likelihood	-18032.43	-17820.85

N=18,139. Dependent variable is Ln(1+assigned prison time in months). Absolute value of groupwise (judge) heteroscedasticity-consistent t-statistics appear in parentheses. Coefficients for Crime and Year dummies not reported in both columns. Coefficients for Judge dummies not reported in column (2).

5 Some Remaining Confounding Influences Addressed

The battery of statistical tests described above suggest that our findings are robust to alternative specifications and incompatible with competing causal mechanisms. Here, we consider three remaining objections. First, one might object that merely controlling for the crime rate does not take into account variation in its effect on the behavior of public officials over the electoral cycle (Levitt 1997). This argument could be developed in two different ways. One might posit that a higher crime rate would lead all judges to raise their sentences as reelection nears in order to satisfy fearful voters. On the other hand, a high crime rate may itself be an artifact of a tendency by liberal judges to coddle criminals, and would therefore yield a smaller proximity effect than comparatively crime-free, conservative districts where punishment is non-controversial.

To address these arguments, we conducted two additional tests. We first re-estimated the model reported in table 5 including the interaction between crime rate and electoral proximity, finding the results nearly unchanged. The electoral proximity coefficient is positive and statistically significant in competitive districts and indistinguishable from zero in retention districts. The estimates suggest no electorally-conditioned effect of crime rates on sentencing. We also re-estimated the earlier model specification dropping the observations from Shawnee County, which had a crime rate fully 34% higher than in the next most dangerous county. Again, the results are nearly unchanged.

A second potential objection is that our results are driven by some intrinsic difference between urban and rural counties. At first glance this seems unlikely because (as we discuss above), the most urban counties are evenly split between retention and competitive systems. Further, in both our regression and matching analyses, we sought to mitigate this potential confounding influence. Nonetheless, we reestimated the models reported in Table 5 separately for the 5 most populous (and urban) districts and the remaining districts. While indications of statistical significance change slightly (in part due to the reduced sample size), we continue to find a larger proximity effect in competitive districts than in retention ones.

For the five largest districts, the proximity effect (in the specification with judge-specific fixed effects) is positive and statistically significant at $p < 0.05$ in the competitive districts and negative and statistically insignificant in retention districts. The estimates using the remaining districts display a familiar pattern, but the coefficient on proximity in the competitive district is significant only at $p < .10$. Nonetheless, we can reject the null hypotheses that the proximity coefficients are identical across selection systems at $p < 0.01$.

Finally, to ensure our results are not driven by peculiarities of particular districts, we re-estimated the main specification 31 times, each time omitting a single district. In all thirty-one cases, the coefficient on electoral proximity is positive and significant in the competitive districts and negative in the retentions districts (but never statistically significant). The average difference between these coefficients is 0.53, with a standard deviation of 0.03. The minimum difference between coefficients is 0.46, about 16% smaller than the observed difference in the model including all observations. In short, we have little reason to believe that these results are due to anything other than the difference in electoral incentives between competitive and retention districts.

19

6   Conclusion

Does the threat of a viable challenger in an election alter the behavior of elected officials and, by extension, the relationship between voters and those officials? This research provides strong evidence that it does. Competitive elections, and the attendant risk of a viable challenger, force incumbent politicians to pay more heed to potential negative voter reactions to their behavior. With respect to this paper's specific object of empirical scrutiny, the risk of challenger entry induces trial judges elected in partisan competitive districts in Kansas to behave more punitively than their peers in that state's retention districts.

Potential challengers might alter incumbent behavior for different reasons. If they choose to run for office, their presence might serve to improve voters' selection of likeminded officials. On the other hand, as we have argued, challengers can also enhance the relationship between voters and incumbents by enhancing the power of electoral incentives. Through their implicit threat to inform voters about the malfeasance of incumbents, for example, challengers may deter that malfeasance in the first place. In our analysis, we find that the sentencing behavior of judges under partisan competitive selection rules is indistinguishable from that of judges under retention rules when election is a far-off prospect, but that the former become more punitive relative to the latter as the electoral threat grows closer. This constitutes empirical confirmation that the increase in the power of incentives caused by the threat of electoral competition dominates the selection effect.

We conclude with some informal observations about the normative implications of these results. The capacity to induce shifts in judicial behavior may not necessarily be an overriding goal in determining the appropriate selection mechanism for lower court judges – or any official for that matter (e.g. Maskin and Tirole 2004). Pandering behavior by elected officials is especially problematic in the presence of severe information asymmetries between them and voters. In the case of trial judges, a normative impulse for consistency in treatment may produce a desire to eliminate institutions that can produce variation in sentencing over time. Likewise, normative arguments concerning the appropriate level of punishment for a particular crime may lead us to favor institutions that produce more or less anticipation and fear by incumbents of punishment at the polls. These are questions we cannot address here. We have sought instead to better identify and understand the extent to which electoral incentives can bind incumbent officials, whether for better or worse.

Appendix A. A Model of Judges' Preferred Sentences

We begin by deriving a judge's induced ideal sentence given a conviction. The judge must sentence $s \in [0, \hat{s}]$, where $\hat{s}$ is the legal maximum sentence. Her ideal sentence for that case is $s_j \in [0, \hat{s})$. The judge has quadratic loss in the deviation from her ideal sentence.

Let $\delta \in (0, 1)$ be the judge's discount factor. The judge receives a (discounted) payoff of one if she wins election and zero if she loses. $T \in \mathbb{R}^+$ represents the time at which electoral pressures are at their maximum, and $t \in [0, T]$ the time elapsed in a judge's term. The probability the judge loses election is equal to $q((\hat{s} - s)/\hat{s})^2$. This is increasing (at an increasing rate) in the distance of s below its maximum $\hat{s}$ (scaled by $\hat{s}$), and $q \in (0, 1)$, a parameter reflecting the sensitivity of negative electoral response to lenient sentencing (which varies across electoral institutions). Thus, if the judge sentences at the maximum ($s = \hat{s}$), she never loses. (Adding to the model a nonzero baseline loss probability would not affect the intuition.) This captures the intuition in the text that judges are threatened electorally by perceived leniency. The squared term captures the notion that the sensitivity of the "fire alarm" increases given more lenient sentencing.

The expected utility to the judge is the sum of the discounted present value of retaining office and the disutility of sentencing away from one's ideal:

$$E[u_j(s; T, t, q, \delta, \alpha, s_j, \hat{s})] = \delta^{T-t}\left(1 - q\left(\frac{\hat{s} - s}{\hat{s}}\right)^2\right) - \alpha(s - s_j)^2,$$

where α captures the judge's tradeoff between implementing preferred sentences and the value of retention.[22] Taking derivatives and solving for s yields an optimal sentence of

$$s^* = \frac{\hat{s}(\delta^{T-t}q + \alpha s_j\hat{s})}{\delta^{T-t}q + \alpha\hat{s}^2}.$$

This is a positive fraction of $\hat{s}$, indicating an interior solution. Next, we turn to comparative statics. We first consider the effect of changing the sensitivity of electoral response:

$$\frac{\partial s^*}{\partial q} = \frac{\alpha\delta^{T-t}\hat{s}^2(\hat{s} - s_j)}{(\delta^{T-t}q + \alpha\hat{s}^2)^2}$$

Because $s_j < \hat{s}$ by assumption, this quantity is always positive. Given the argument that competitiveness implies greater potential adverse electoral consequences for leniency, the empirical implication is that ceteris paribus, sentencing should be higher in partisan competitive districts than retention districts. Second, we consider the electoral proximity effect:

$$\frac{\partial s^*}{\partial t} = \frac{-\alpha\delta^{T-t}\ln(\delta)q\hat{s}^2(\hat{s} - s_j)}{(\delta^{T-t}q + \alpha\hat{s}^2)^2}$$

---

[22]Note that α captures in reduced form any expectations by the judge that he or she will experience psychological costs of deviations from his or her ideal not just immediately but over time as well. Critically, however, these costs are uncorrelated with the judge's position in the electoral cycle.

This is also always positive (recall that $\ln(\delta)$ is negative), indicating that all else equal, sentencing should become more punitive as elections draw closer. Finally, the interaction between sensitivity and proximity is indicated by the cross-partial derivative:

$$\frac{\partial^2 s^*}{\partial t \partial q} = \frac{-\alpha \hat{s}^2 (\hat{s} - s_j) \ln(\delta) \delta^{T-t}}{(\delta^{T-t} q + \alpha \hat{s}^2)^3} \left(\alpha \hat{s}^2 - \delta^{T-t} q \right)$$

This can be *either* positive or negative. If $\hat{s}$ is greater than $\sqrt{\delta^{T-t} q / \alpha}$, then we should expect the electoral proximity effect to be enhanced in partisan competitive districts. If $\hat{s}$ lies above this quantity, we should expect the proximity effect to be larger in retention districts. Only at the knife's edge should we expect no difference in proximity effect across systems.

Because prosecutors and defendants negotiate plea bargains in the shadow of the judge, who both has the power to reject them as well as to sentence given a conviction following a failure to reach an agreement, both negotiated pleas and sentencing at trial will reflect the judge's preferred sentence (LaCasse and Payne 1999) and be similarly responsive to changes in parameter values.

Appendix B. Constructing the Electoral Proximity Measure

In order to derive a measure that is comparable across different selection mechanisms, this variable must capture several relevant features of the electoral calendar. In competitive systems, an incumbent judge might face 1) a challenger in the primary but not the general election; 2) a challenger in the general election but not the primary; 3) a challenger in both the primary and the general; or 4) no challenger. Given these possibilities, there are three relevant dates for an incumbent judge j in election cycle c: the filing deadline, labeled τ_c^f, the primary election day, τ_c^p, and the general election day, τ_c^g. We first calculate a "base" measure of electoral proximity at time t: the fraction of the judge's four year term elapsed before the next filing deadline, $\frac{\tau_c^f - t}{\tau_c^f - \tau_{c-1}^f}$. When the filing deadline is imminent, this score is equal to one. In case 1, when the judge only faces a challenger in the primary, the measure remains at one until the primary date. In cases 2 and 3, it remains at one until the general election date. In case 4, it drops immediately to the base rate after the deadline.

Formally, let ρ_{jct} be an indicator variable equal to one if and only if judge j is running for reelection at the end of cycle c as of time t. (ρ_{jct} may equal zero if the judge retired or if he or she lost either the primary or general election.) Let θ_{jc}^p be an indicator variable equal to one if and only if judge j faces a primary challenger in election cycle c, and θ_{jc}^g a corresponding indicator for a general election challenger. Electoral proximity for a judge in a partisan competitive district is

$$PROX_{jct}^P = \begin{cases} 0 & \text{if } \rho_{jct} = 0; \\ 1 & \text{if } \tau_c^f < t < \tau_c^p, \rho_{jct} = 1, \text{ and } \theta_{jc}^p = 1; \\ & \text{or } \tau_c^f < t < \tau_c^g, \rho_{jct} = 1, \text{ and } \theta_{jc}^g = 1; \\ \frac{\tau_c^f - t}{\tau_c^f - \tau_{c-1}^f} & \text{otherwise.} \end{cases}$$

Because of the way electoral competition is coded in the competitive systems, judges who face challengers in the general election have an electoral proximity score of one between the filing date and the general election. For comparability, in retention systems we code electoral proximity for

the period prior to the filing date as the number of days elapsed since the last filing date divided by the total number of days in a term. After the filing date, if an incumbent judge has chosen to run for retention, the electoral proximity variable remains at one until the day after the general election and is then coded as before. If she does not choose to run or is defeated, the electoral proximity score remains at zero.[23] Formally,

$$PROX_{jct}^R = \begin{cases} \frac{\tau_c^f - t}{\tau_c^f - \tau_{c-1}^f} & \text{if } \tau_{c-1}^f < t < \tau_c^f; \\ 1 & \text{if } \tau_c^f < t < \tau_c^g \text{ and } \rho_{jct} = 1; \\ 0 & \text{otherwise.} \end{cases}$$

Appendix C. Derivation of the Maximum Likelihood Estimator

To operationalize proportionality, we assume the induced ideal punishment of judge j for convicted defendant i at time t is a multiplicative function of the defendant's *culpability* and the judge's *punitiveness*. Culpability $c_{it} \in \mathbb{R}^+$ refers to a set of circumstances or a fact pattern associated with the commission of a crime, including the criminal history of the defendant, the nature of the crime itself, and victim characteristics. Punitiveness, $a_{ijt} \in \mathbb{R}^+$ can emerge from several sources: judge-specific time invariant characteristics such as philosophy or ideology, the position of the judge in his or her electoral calendar, or any possibly discriminatory motivations associated with specific defendant characteristics.

The utility to judge j of sentence s for defendant i at time t is single-peaked and given by $u_j(s; c_{it}, a_{ijt}) = -g(|a_{ijt} c_{it} - s|)$, where $g(\cdot)$ is an arbitrary increasing function. The judge's unconstrained preferred sentence is $s_{ijt}^* = a_{ijt} c_{it}$. We model both punitiveness and culpability as exponential functions of observable and unobservable (to the analyst) features of the judge, defendant, crime, etc.: $a_{ijt} = \exp(X_{ijt}' \beta + \varepsilon_{ijt}^a)$; and $c_{it} = \exp(Z_{it}' \gamma + \varepsilon_{it}^c)$. For each judge, we assume ε_{ijt}^a and ε_{it}^c are distributed multivariate normal with mean vector $\mathbf{0}$ and covariance matrix Σ_j. Substituting and taking logs gives

$$\ln(s_{ijt}^*) = X_{ijt}' \beta + Z_{it}' \gamma + \varepsilon_{ijt}^a + \varepsilon_{it}^c. \tag{1}$$

In principle, β and γ could be estimated via least squares (although separate constant terms would not be identified), with a standard error adjustment to account for judge-specific groupwise heteroscedasticity. Two difficulties persist. First, statutory maximum sentences limit how long a sentence a judge can assign. In those cases, we treat the data as right-censored: the judge may have wanted to impose a larger sentence, but was unable to.

A more pervasive problem emerges because a sentence can consist of two components: the *incarcerative* portion, consisting of jail time, and a non-incarcerative portion, consisting of, for

---

[23]Our results do not change if we simply drop judges who chose not to run again. Including them, however, is a conservative strategy that biases against detecting a proximity effect in either system. Doing so dilutes the sample of those who are bound by electoral incentives with those who are not. With this scaling, we effectively assume that there is a shared "election window" between the filing date and the general election for all retention judges seeking retention and those judge in competitive systems who will face a challenger. (Simply coding electoral proximity in the retention districts as the elapsed fraction of the term produces substantively identical results.)

example, fines, probation, and community service. Difficulties emerge because we cannot place non-incarcerative terms on the same metric as jail or prison time, and because a majority of cases involve zero prison time.

We recast the scaling problem as a left-censoring problem by assuming that from the perspective of the defendant, the worst non-incarcerative sentence is preferable to the most lenient prison term.[24] Let γ be the most punitive sentence short of imprisonment. Representing the prison portion of sentence s_{ijt}^* as p_{ijt}, the dependent variable becomes $\ln(p_{ijt} + \gamma)$. As noted in the text, we set $\gamma = 1$, an assumption that implies one month in prison is exactly twice as painful as the maximum non-incarcerative sentence. Substituting values for γ between 0.1 and 10 has no influence on our substantive results. When a sentence involves zero prison time, it is left-censored at $\ln(1) = 0$. The right-censoring implied by the statutory maximum sentences and the left-censoring implied by non-incarcerative sentences yields the two-limit Tobit likelihood function (with groupwise heteroscedasticity adjustment) in which the right-censoring point differs from observation to observation.

---

[24]Given the felonies to which we restrict our attention, this is not particularly controversial. If this assumption were false, convicts would, when offered terms of probation, turn them down in favor of prison time. We know of no instance in which this has occurred.

References

Alvarez, R. Michael. 1997. *Information and Elections.* Ann Arbor, MI: University of Michigan Press.

Arnold, R. Douglas. 2004. *Congress, the Press, and Political Accountability.* Princeton, NJ: Princeton University Press and Russell Sage Foundation.

Arnold, R. Douglas. 1993. "Can Inattentive Citizens Control Their Elected Representatives?" in Lawrence C. Dodd and Bruce I. Oppenheimer (eds.), *Congress Reconsidered,* 5th ed. (Washington: CQ Press), pp. 401-416.

Barro, Robert. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14: 19-42.

Bartels, Larry M. 1991. "Constituency Opinion and Congressional Policy Making: The Reagan Defense Buildup." *American Political Science Review* 85(2): 457-474.

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

Brady, David W., Brandice Canes-Wrone, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members Voting." *American Political Science Review* 96: 127-140.

Cox, Gary W. and Jonathan N. Katz. 1996. "Why Did the Incumbency Advantage in U.S. House Elections Grow?" *American Journal of Political Science,* 40: 478-97.

Diamond, Alexis, and Jasjeet S. Sekhon. 2005. "Genetic Matching for Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." Mimeo, Survey Research Center, University of California at Berkeley.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation,* eds. Bernard Manin, Adam Przeworski, and Susan Stokes. New York: Cambridge University Press, 55-97.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50: 5-25.

Gordon, Sanford Clark, Gregory Huber, and Dimitri Landa. 2006. "Challenger Entry and Voter Learning." New York University, Department of Politics Mimeo.

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: 247-263.

LaCasse, Chantale, and Abigail A. Payne. 1999. "Federal Sentencing Guidelines and Mandatory Minimum Sentences: Do Defendants Bargain in the Shadow of the Judge?" *Journal of Law & Economics,* 42: 245-69.

Lee, David S., Enrico Moretti, and Matthew J. Butler. 2004. "Do Voters Affect of Elect Politices? Evidence from the U.S. House." *Quarterly Journal of Economics* 119: 807-859

Levitt, Steven D. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review* 87: 270-290.

Maskin, Eric and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government" *American Economic Review* 94: 1034-1054.

Mathias, Sara. 1990. *Electing Justice: A Handbook of Judicial Election Reforms.* Chicago: American Judicature Society.

Mayhew, David R. 1974. *Congress: The Electoral Connection.* New Haven: Yale University Press.

McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28: 165-179.

Miller, Warren and Donald Stokes. 1963. "Constituency influence in Congress." *American. Political Science Review* 57: 45-56.

Roberts, Julian V., and Don Edwards. 1989. "Contextual Effects in Judgments of Crimes, Criminals, and the Purposes of Sentencing." *Journal of Applied Pysychology* 19: 902-917.

Sekhon, Jasjeet S. 2005. "Matching: Multivariate and Propensity Score Matching Software with Automated Balance Search." Available at http://jsekhon.fas.harvard.edu/matching/.

Sheldon, Charles H., and Nicholas P. Lovrich. 1983. "Knowledge and Judicial Voting: The Oregon and Washington Experience." *Judicature* 67: 234-45.

Warr, Mark. 1995. "Poll Trends: Public Opinion on Crime and Punishment" *Public Opinion Quarterly* 59: 296-310.

Watson, Richard and Rondal Downing. 1969. *The Politics of the Bench and the Bar.* New York: John Wiley and Sons, Inc.